



# Epidémiologie et génétique humaine de l'ulcère de Buruli

Quentin Vincent

## ► To cite this version:

Quentin Vincent. Epidémiologie et génétique humaine de l'ulcère de Buruli. Génétique. Université René Descartes - Paris V, 2014. Français. NNT : 2014PA05S019 . tel-01249580

**HAL Id: tel-01249580**

**<https://theses.hal.science/tel-01249580>**

Submitted on 4 Jan 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSITE PARIS DESCARTES**

**ECOLE DOCTORALE 420 SANTE PUBLIQUE**

Année 2014

N°

**THESE**

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE PARIS DESCARTES**

Discipline : Génétique Statistique

Ecole Doctorale ED420 de Santé Publique

Présentée et soutenue publiquement le 28 novembre 2014

Par Quentin VINCENT

Né le 30/11/1981 à Paris

**Epidémiologie et Génétique Humaine**  
**de l'ulcère de Buruli**

Jury : Dr Florence Demenais (Président)

Dr Jean-François Guégan (Rapporteur)

Pr Bertram Müller-Myhsok (Rapporteur)

Dr Eric Comte (Examineur)

Dr Etienne Patin (Examineur)

Pr Bernard Carbonnelle (Membre d'honneur)

Dr Alexandre Alcaïs (Directeur de thèse)



# Epidémiologie et Génétique Humaine de l'ulcère de Buruli

---

## Sommaire

Table des illustrations.....	4
Remerciements .....	9
Résumé .....	13
Summary .....	14
Liste de Publications.....	15

## Introduction

Etat des connaissances.....	21
L'ulcère de Buruli : perspective historique.....	21
Epidémiologie de l'infection à <i>M. ulcerans</i> .....	25
L'infection à <i>M. ulcerans</i> , une maladie vectorielle ? .....	31
Aspects cliniques : diagnostic, évolution, séquelles.....	37
Diagnostic et prise en charge .....	43
Diagnostic positif et diagnostic différentiel .....	43
Prise en charge .....	46
Physiopathologie de l'infection à <i>M. ulcerans</i> .....	51
L'analyse du génome de <i>M. ulcerans</i> révèle une évolution réductrice .....	51
La mycolactone : facteur de virulence majeur de <i>M. ulcerans</i> .....	54
<i>M. ulcerans</i> : un germe intracellulaire ou extracellulaire ?.....	56
Exploration de l'épidémiologie et de la génétique humaine de l'ulcère de Buruli : motivations et objectifs .....	61

## Epidémiologie de l'ulcère de Buruli

Epidémiologie de l'UB : contexte .....	69
Epidémiologie de l'UB : population et méthodes .....	71
Epidémiologie de l'UB : résultats .....	75
Diagnostic et prise en charge de l'UB au CDTUB de Pobè .....	75
La présentation clinique la plus classique de l'UB est une ulcération large d'un membre inférieur	77



Le sexe masculin est surreprésenté parmi les cas pédiatriques d'UB tandis que le sexe féminin est surreprésenté parmi les cas adultes d'UB.....	79
La présentation clinique de l'UB varie en fonction de l'âge et du sexe .....	81
Certaines présentations cliniques concentrent le risque de séquelles fonctionnelles et définissent la sévérité de l'UB.....	83
L'infection par le VIH est associée à la sévérité de la présentation clinique de l'UB.....	86
Epidémiologie de l'UB : discussion.....	87

### **Génétique humaine de l'ulcère de Buruli : aspects mendéliens**

Génétique mendélienne de l'UB : introduction .....	97
Génétique mendélienne de l'UB : sujets et méthodes .....	98
Génétique mendélienne de l'UB : résultats .....	102
Cas clinique.....	102
Investigations génétiques.....	105
Génétique mendélienne de l'UB : discussion.....	110

### **Génétique humaine de l'ulcère de Buruli : aspects complexes**

Collection des données cliniques et biologiques .....	121
Etude de faisabilité.....	121
Etude de faisabilité globale : voyages de terrain .....	121
Etude de faisabilité logistique : du prélèvement sanguin au génotypage .....	125
Recrutement.....	129
Recrutement prospectif et rétrospectif des cas et des témoins .....	129
Phénotype : Confirmation diagnostique des cas et des témoins.....	131
Gestion de la collection biologique .....	132
Génotypage de la cohorte primaire .....	134
Description de l'échantillon primaire .....	134
Puce de génotypage .....	134
Contrôle qualité.....	136
Contrôle qualité par individu.....	136
Contrôle qualité par SNP .....	139
Imputation de génotypes additionnels .....	142
Structure génétique de la cohorte primaire .....	145
Méthodes d'identification.....	146
Structure génétique de la cohorte de Pobè .....	151

Liens familiaux cryptiques .....	151
Structure de population à l'échelle mondiale .....	152
Structure de population à l'échelle locale.....	154
Analyse d'association pangénomique .....	157
Méthodes d'association .....	157
Modèles génétiques : codage du génotype .....	157
Modélisation du lien phénotype-génotype.....	161
Analyse de la cohorte primaire .....	169
Résultats préliminaires.....	169
Prise en compte de la structure génétique de la cohorte primaire : choix du nombre de composantes principales d'ajustement .....	171
Prise en compte de la structure génétique de la cohorte primaire : ajustement sur composantes principales et modèle mixte .....	174
Corrélation entre analyses .....	177
Pvalues minimales : Manhattan plot et meilleurs signaux.....	180
Association : cas des SNPs imputés.....	183
Réplication des signaux de la cohorte primaire .....	185
Sélection des SNPs pour l'étude de réplication .....	185
Cohorte de réplication.....	188
Etude d'association pangénomique : discussion.....	189
 <b><u>Conclusion</u></b>	
Conclusion .....	199
Bibliographie.....	205
Annexes .....	225

## Table des illustrations

Figure 1. Nombre d'articles référencés dans PubMed pour <i>M. tuberculosis</i> , <i>M. leprae</i> et <i>M. ulcerans</i> .....	18
Figure 2. Descriptions cliniques originales de l'UB.....	22
Figure 3. Epidémiologie de l'infection à <i>M. ulcerans</i> .....	26
Figure 4. Transmission de <i>M. ulcerans</i> .....	32
Figure 5. Formes pré-ulcératives de l'infection à <i>M. ulcerans</i> .....	36
Figure 6. Formes ulcérées, complications et séquelles.....	38
Figure 7. Réactions paradoxales.....	40
Figure 8. Confirmation diagnostique - prélèvements .....	43
Figure 9. Aspects anatomopathologiques de l'infection à <i>M. ulcerans</i> .....	44
Figure 10. <i>M. ulcerans</i> est proche de <i>M. marinum</i> mais a subi une évolution réductrice.....	52
Figure 11. La mycolactone, toxine diffusible cytotoxique et immunomodulatrice de <i>M. ulcerans</i> .....	53
Figure 12. Modèle physiopathologique de l'infection à <i>M. ulcerans</i> .....	58
Figure 13. Du contact avec <i>M. ulcerans</i> au développement de signes cliniques :.....	60
Figure 14. Schéma de l'étude .....	62
Figure 15. Le lieu de l'étude : Pobè, région de Ouémé-Plateau, Bénin.....	64
Figure 16. Distribution des cas d'UB en fonction de l'année du diagnostic (A), du mois du diagnostic (B) et du statut de confirmation par examen complémentaire (C).....	74
Figure 17. Présentation clinique de l'UB : forme clinique, site atteint, nombre et taille des lésions.....	76
Figure 18. Présentation clinique de l'UB : impact de l'âge au diagnostic .....	78
Figure 19. Survenue de séquelles à long-terme chez les patients en fonction de la sévérité de la présentation clinique de l'UB .....	84
Figure 20. Sévérité de la présentation clinique de l'UB en fonction du statut VIH.....	85
Figure 21. Arbre généalogique et phénotype clinique de la famille BU-A.....	100
Figure 22. Evolution clinique du patient index – 2005-2010 .....	101
Figure 23. Résultats de l'analyse de liaison par cartographie d'homozygotie .....	104
Figure 24. Résultats de l'analyse des variations en nombre de copies - chromosome 8.....	108
Figure 25. Structure et fonction des défensines .....	112
Figure 26. Le locus des défensines en 8p23 .....	113
Figure 27. Impact du délai d'extraction et de la température de conservation sur le rendement et l'intégrité de l'ADN : conditions de laboratoire .....	124
Figure 28. Impact du délai d'extraction et de la température de conservation sur le rendement et l'intégrité de l'ADN : conditions de terrain .....	127
Figure 29. Bases de données disponibles et assignation finale du statut cas-témoin au moment du génotypage de la cohorte primaire.....	130
Figure 30. Alignement des noms et prénoms par l'algorithme BLAST.....	132
Figure 31. Contrôle qualité par individu, n=816.....	135
Figure 32. Contrôle qualité par SNP, 2374209 SNPs, 803 individus.....	138
Figure 33. Principe de l'assignation de génotypes par puce à ADN.....	139
Figure 34. Principe de l'imputation de génotypes .....	141

Figure 35. Structure génétique de la cohorte : identité par état et ACP à l'échelle mondiale .....	149
Figure 36. Structure génétique de la cohorte : ACP à l'échelle locale .....	153
Figure 37. Vue aérienne de Manhattan : analyses pangénomiques d'association de l'UB .....	168
Figure 38. Choix du nombre de composantes principales à inclure pour ajustement .....	170
Figure 39. Intersections des 100 signaux les plus significatifs pour chaque analyse .....	176
Figure 40. Manhattan plot des pvalues minimales : SNPs géotypés.....	178
Figure 41. Manhattan plot des pvalues minimales : SNPs imputés .....	181
Figure 42. Mode de sélection des 384 SNPs géotypés dans la cohorte de réplication.....	184
Figure 43. Manhattan plot : SNPs sélectionnés pour la réplication .....	186



« Non esistono proprie montagne, si sa,  
esistono però proprie esperienze.  
Sulle montagne possono salirci molti altri,  
ma nessuno potrà mai invadere le esperienze  
che sono e rimangono nostre. »  
Walter Bonatti

« Comment y arrive-t-on ?  
C'est l'action de se frayer un passage  
à travers un mur de fer invisible,  
qui semble se trouver entre ce que l'on *sent*,  
et ce que l'on *peut*. »  
Vincent Van Gogh

« J'ai des doutes sur le changement d'heure en été, (...)  
des réserves quant à la question d'angle pour le canapé, (...)  
sur la remise à flot de la crème renversée, (...)  
j'ai des doutes, est-ce que vous en avez ? »  
Alain Bashung

« C'est marrant,  
quand on avance,  
on dirait que la terre  
elle vient avec nous. »  
Un enfant

« Tout dans ce monde,  
qui est le seul où nous ayons accès,  
est plus étrange  
que l'habitude ne nous porte à le croire. »  
Zénon in *L'œuvre au noir*  
Marguerite Yourcenar



## Remerciements

Avant tout, j'aimerais remercier l'équipe du CDTUB de Pobè qui s'est lancée dans notre aventure, à fond, et à l'aveugle. J'aimerais donc remercier Annick Chauty, Marie-Françoise Ardant et Pierre Ardant. J'aimerais aussi remercier chaque membre du personnel du CDTUB. Tout était si ténu ! J'ai passé en tout et pour tout deux ou trois petites semaines à Pobè réparties en deux séjours en juin 2010 et juin 2011. Immédiatement, j'aimerais remercier Laurent Marsollier, notre tête de pont... Laurent a instauré la confiance, je crois, en prenant à sa charge, semaine après semaine, le diagnostic PCR à Angers des suspicions d'ulcère de Buruli vus à Pobè. Incroyable et fidèle collaboration transcontinentale, qui dure depuis 2006. Encore plus incroyable peut-être, sa volonté et celle d'Estelle Marion d'équiper, de former et d'accompagner l'équipe de biologie médicale de Pobè pour la réalisation de cet examen sur place, depuis deux ans environ. A mon sens, c'est cette confiance construite sur l'activité de soin qui a rendu possible d'envisager à Pobè un projet si étrange et étranger que l'étude de la génétique *humaine* de la réponse à *Mycobacterium ulcerans*. Merci à Laurent et à son équipe donc, d'avoir accompli cette médiation initiale.

Merci à Annick, en tant que directrice médicale, d'avoir écouté et accepté notre projet. Merci à Annick, en tant qu'hôte, de nous avoir accueillis à Pobè alors que l'activité ne faiblissait pas. Merci à Marie-Françoise et à Annick d'avoir répondu à mes questions régulières par email sur l'ulcère de Buruli et sur leur activité à Pobè. Ces échanges m'ont permis de comprendre au mieux la nature des données qui étaient à ma disposition afin de pondérer le sûr et l'incertain pour construire une présentation qui fasse honneur, peut-être, aux heures et aux années passées à soigner les patients au cœur de la campagne africaine. Merci à Marie-Françoise et à Pierre (et au ciment Lafarge !) de nous avoir hébergés à Onigbolo. J'ai ainsi été initié conjointement à la culture du Ouémé-Plateau (grandeur et misère du royaume de Kétou ; mythes et réalités des cultes vaudous ; le sodabi, entre plaisir et douleur...) et à la transformation du calcaire (je n'ai malheureusement pas réussi à inclure une photo des hauts-fourneaux dans ma thèse)...



Je remercie également Laurent et son équipe pour leur accueil à Angers. Après la phase de mise au point initiale, les extractions d'ADN ont été réalisées chaque semaine à Angers, d'abord par Camille Dantec, qui a su faire face certaines semaines aux envois massifs des journées de recrutement de cas rétrospectifs, puis par Jérémie Babonneau. Nous avons passé ensuite plusieurs semaines avec Jérémie à prendre soin de ces précieux échantillons, les doser, les normaliser, les redoser, les randomiser... Grande angoisse pour moi que ces semaines à Angers où je me retrouvais dans le rôle de l'expérimentaliste de paillasser. Merci Jérémie pour ta confiance et ta patience lors de ces longues parties de bataille navale! Merci à Estelle pour les soirées, à l'ambiance tantôt moulignée, tantôt orageuse et tantôt révélatrice... Merci à Guillaume Vogt, Maya Chrabieh, Emmanuelle Jouanguy, Erwin Schurr et Julien Guernon pour leurs conseils sur la gestion de la collection biologique. Il s'avèrera en octobre 2014 que je ne sais toujours pas correctement fermer des plaques 96 puits... Mais pourtant le GWAS est là.

Merci aux membres de l'U550 devenue U980 puis U1163 de m'avoir écouté lors de mes topos du vendredi. Autre grande angoisse de ma thèse que ces mises à nues tous les trois ou quatre mois devant Paris et New York, devant un auditoire essentiellement expérimentaliste. Il est étonnant de pouvoir aborder un même thème, la génétique humaine des maladies infectieuses, par des voies si différentes. Il y a une certaine beauté à voir les expérimentalistes compléter élégamment des puzzles, et à nous voir, épidémiologistes, nager le papillon dans un océan de corrélations. Merci à Jean-Laurent Casanova, Laurent Abel et Alexandre Alcaïs de m'avoir donné une place au laboratoire et de m'avoir accompagné dans mes questionnements, parfois légitimes, parfois involontairement déplacés, depuis maintenant plus de dix ans. Avec les anciens du laboratoire, vous avez été les témoins de nombreux choix majeurs de ma vie, comme entreprendre des études de médecine après l'ENS et comme passer de la pédiatrie à la santé publique. Merci de votre souplesse à mon égard. Merci de m'avoir permis de tant voyager au cours de ma thèse, sur le terrain ou en congrès, au Bénin, au Vietnam, en Inde, au Canada, aux Etats-Unis, en Suisse. Ces temps de respiration, entre rencontres et solitude, sont de précieux nourriciers pour moi. Merci aux oreilles qui m'ont écouté et

questionné, Anne Puel, Emmanuelle Jouanguy, Alexandre Alcaïs, Jean Gaschignard, Erika Della Mina, Audrey Grant, Aziz Belkadi, Jérôme Flatot. Merci à tous ceux qui favorisent un lien festif dans le groupe. Je pense en particulier à Vanessa Sancho-Shimizu, Laura Israël, Francesca Conti, Cécile Fligny, Antoine Guérin.

Merci à tous les membres présents et passés de l'équipe épidémio pour l'expérience partagée : Laurent Abel, Alexandre Alcaïs, Sabine Plancoulaine, Aurélie Cobat, Jérôme Flatot, Etienne Patin, Vincent Pedergrana, Ghislain Grange, Jean Gaschignard, Kristel Cosker, Aziz Belkadi, Laurène Syx, Fabienne Jabot-Hanin, Frédégonde About. Merci également à Jean-Philippe Jaïs. Merci à Cédric Laouénan, un autre ancien de l'équipe épidémio, pour nos discussions. Merci à Patrick Nitschké, de la plateforme bioinformatique, pour nos nombreuses interactions aux débuts de Polyweb. Merci à l'équipe de l'U946 chez qui j'ai passé six mois d'internat juste avant mon arrivée à l'unité, en particulier à Florence Demenais et Emmanuelle Bouzigon. C'est une grande chance que d'avoir pu vivre plusieurs mois dans un autre laboratoire de génétique épidémiologique parisien avant de venir à l'U550. Merci à ceux qui m'ont fait confiance pour donner des cours et des formations, Patrick Langue, Alexandre Alcaïs, Hervé Perdry, Anne-Louise Leutenegger, Emmanuelle Génin, Lluís Quintana-Murci. Merci aux compagnons de congrès et de voyage pour ces temps hors du temps, Hughes Aschard, Jérémie Manry, Vinicius Fava, Marianna Orlova, la soupe Mi, Erwin Schurr, Aurélie Cobat, Marcello Mira, Steven Gazal, Estelle Marion.

Merci aux membres de mon jury de thèse d'avoir bien voulu lire et évaluer mon travail. Florence Demenais préside ce jury. Merci à elle de m'avoir accueilli à l'U946 en 2009. Son enthousiasme constant pour la discipline me marque, m'étonne et me questionne. A la vie noctambule... Dans l'écho des voix *blues* à Chicago ! Merci à Jean-François Guégan et à Bertram Müller-Myhsok d'avoir accepté d'être rapporteurs de ma thèse et d'avoir toléré quelques écarts calendaires. Merci à eux pour leurs observations sur mon manuscrit initial. Merci à Eric Comte d'avoir accepté d'être examinateur de ma thèse. Je suis profondément heureux que Médecins Sans

Frontières soit ainsi représenté au sein de mon jury. Merci à Etienne Patin d'avoir accepté d'être examinateur de ma thèse. Rien en biologie n'a de sens sinon à la lumière de l'évolution dit-on, ce qui, d'une pierre deux coups, m'explique peut-être sa vocation de chercheur professionnel et d'éleveur amateur ! Merci à Bernard Carbonnelle d'avoir d'abord accepté d'être examinateur de ma thèse et d'avoir ensuite accepté de bonne grâce les tracasseries administratives qui l'ont promu au grade de Membre d'Honneur de ce jury !

Merci à Alexandre Alcaïs, le scientifique, d'avoir dirigé ma thèse avec constance, anticipation, fermeté et régularité, suivant sans faillir le précepte d'un sage des temps anciens qui avait dit : « Si tu t'y mets six heures avant la deadline, tu n'y passeras pas plus de six heures ». Merci à Alexandre Alcaïs, l'ami, pour le commerce des âmes.

Merci à ma famille et à mes amis qui tentent de suivre sans trop de terreur mon parcours hors-piste. Merci à mes guides de haute montagne, Mathieu, Rachel, Mylène et Autre. Merci à Delphine, ma cousine, et à Benjamin, mon pote, de leur confiance inaltérée, qui, moi-même, me dépasse. Merci aux enfants, en particulier à mes neveux, ma nièce, et mes filleuls, Julien, Maxence, Pauline, Camille, Romain et Timothée et Thomas d'exprimer si librement leur énergie de vie. A chaque instant, l'instant.

Merci à l'association Raoul Follereau France, en particulier à Bénédicte de Charrette et à Christian Johnson, pour leur soutien à notre projet. Merci à Didier Agossadou, coordonnateur du Plan National de Lutte contre l'Ulcère de Buruli au Bénin, de nous avoir soutenu. Merci à la Fondation pour la Recherche Médicale et à l'Agence Nationale de la Recherche pour les financements qui ont permis de mener à bien ce travail. Merci à l'Institut Imagine de m'avoir accordé la bourse Santé-Sciences qui a financé ma thèse. Merci à Guillaume Huart et Emilie Resweber-Gouin de m'avoir accueilli dans Imagine avant même Imagine.

Merci à chaque patient d'avoir accepté de participer à notre étude.

## Résumé

L'ulcère de Buruli (UB), infection à *Mycobacterium ulcerans*, troisième mycobactériose mondiale, connaît une émergence rapide depuis 1980, essentiellement dans les pays d'Afrique subsaharienne. Jusqu'ici, les connaissances épidémiologiques sur l'UB étaient fondées sur des séries de cas cliniques non confirmés par laboratoire. Nous avons constitué la plus grande cohorte de cas confirmés à ce jour rassemblant plus de 1200 patients traités au CDTUB de Pobè au Bénin entre 2005 et 2011, afin de décrire l'épidémiologie clinique de la maladie et d'explorer l'architecture génétique de la susceptibilité à cette maladie.

Les patients atteints d'UB sont des enfants (âge médian au diagnostic de 12 ans), présentant une lésion unique (96%), large (plus de 15 cm, 36%), ulcéral (66%) du membre inférieur (60%). Nous rapportons une présentation clinique atypique de l'UB, dans laquelle les patients présentent exclusivement une ostéomyélite à *M. ulcerans*. Le sex-ratio varie avec l'âge : les garçons sont majoritaires parmi les enfants (57% de patients masculins chez les moins de 15 ans), et les femmes parmi les adultes (33% de patients masculins). La présentation clinique dépend de l'âge et du sexe. 9% des patients masculins ont présenté une ostéomyélite contre 4% des patients féminins. Un an après la fin du traitement, 22% des patients présentent des séquelles fonctionnelles fixées. Une présentation clinique comportant une lésion oedémateuse, osseuse, de grande taille ou plusieurs lésions est significativement associée avec le développement de séquelles fonctionnelles (OR 7.64, IC95% [5.29-11.31]). Les patients coinfectés par le VIH ont un risque significativement plus élevé de développer un UB sévère (OR 2.77, IC95% [1.32-6.33]).

Nous avons exploré l'architecture génétique de la susceptibilité à l'UB dans une perspective mendélienne et une perspective complexe. Le cas le plus sévère de la maladie observé dans ce centre appartient à une famille consanguine dans laquelle la ségrégation du phénotype suggère un défaut génétique mendélien récessif. Une analyse de liaison génétique par cartographie d'homozygotie suggère l'implication du locus des bêta-défensines sur le chromosome 8 dans la pathogénèse de l'UB, et mène à l'identification d'une délétion homozygote ségrégeant parfaitement avec la maladie. Dans une perspective complexe, une étude d'association pangénomique a été réalisée après génotypage d'une cohorte de 400 cas et 400 témoins exposés sur plus de 2 millions de SNPs par la puce Illumina Omni2.5 et a permis l'identification de nombreux signaux d'intérêt. L'étude de réplication est en cours.

La compréhension de la physiopathologie de l'infection à *M. ulcerans* est cruciale pour générer de nouvelles pistes thérapeutiques et vaccinales. La dissection du contrôle génétique de l'infection par l'hôte est en ce sens indispensable.

### MOTS-CLES

Maladie infectieuse, Epidémiologie, Génétique Humaine Mendélienne et Complexe, Mycobactérie, *Mycobacterium ulcerans*, Ulcère de Buruli, Bénin, Analyse de liaison, Défensine, Analyse d'association pangénomique, GWAS.

## Summary

Buruli ulcer (BU), caused by *Mycobacterium ulcerans*, is the third most frequent mycobacteriosis worldwide. It has been rapidly emerging in sub-Saharan African countries since 1980. Until now, knowledge of BU epidemiology relied on series of non laboratory-confirmed clinical cases. From 2005-2011, we recruited the current largest cohort of laboratory-confirmed cases (more than 1,200 patients) at the Pobe CDTUB, Benin, to describe the clinical epidemiology of the disease and to explore the genetic architecture of human susceptibility to BU.

Typically, patients with BU were children (median age at diagnosis 12 years) presenting with a unique (96%) large ( $\geq 15$  cm, 36%) ulcerative (66%) lesion of the lower limb (60%). Atypical clinical presentation of BU included osteomyelitis with no identifiable present or past BU skin lesions. The sex ratio of BU widely varied with age, with male patients accounting for 57% of patients aged 15 years and younger, but only 33% of those older than 15 years. Clinical presentation of BU was significantly dependent on age and sex. 9% male patients had BU osteomyelitis, whereas only 4% of female patients did. 1 year after treatment, 22% of patients with follow-up information presented with permanent functional sequelae. Presentation with oedema, osteomyelitis, or large ( $\geq 15$  cm in diameter), or multifocal lesions was significantly associated with occurrence of permanent functional sequelae (OR 7.64, 95% CI 5.29–11.31) and operationally defines severe BU. When coinfecting with HIV, patients had a significantly higher risk to develop severe BU (OR 2.77, IC95% [1.32-6.33]).

We explored the genetic architecture of susceptibility to BU in both mendelian and complex genetic frameworks. The most severe case of the disease to have been treated at the Pobe CDTUB belonged to a consanguineous family in which the segregation of the phenotype was indicative of a recessive mendelian genetic defect. Genetic linkage analysis by homozygosity mapping suggested the implication of the beta-defensin locus on chromosome 8 in BU pathogenesis and led to the identification of a homozygous deletion, which co-segregated perfectly with the disease in the family. In a complex genetics approach, we undertook a genome-wide association study, which involved the genotyping of more than 2 million SNPs (Illumina Omni2.5) in a cohort of 400 cases and 400 exposed controls. We identified many signals of interest. The replication study is ongoing.

Understanding BU physiopathology is crucial to the development of efficient vaccines and drugs. Dissection of the genetic control of the infection by *M. ulcerans* by its human host therefore constitutes an indispensable step.

## KEYWORDS

Infectious disease, Epidemiology, Complex and Mendelian Human Genetic, Mycobacteriosis, *Mycobacterium ulcerans*, Buruli ulcer, Benin, Linkage Analysis, Defensin, Genome-Wide Association Study.

## Liste de Publications

### Publications issues directement du travail de thèse :

**Vincent QB**, Ardant M-F, Adeye A, ...//..., Chauty A, Alcaïs A. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health*. jul 2014;2(7):e422-e430.

**Vincent QB**, Ardant M-F, Marsollier L, Chauty A, Alcaïs A. HIV infection and Buruli ulcer in Africa. *Lancet Infect Dis*. September 2014;14(9):796-7.

Pommelet V, **Vincent QB**, Ardant M-F, ...//..., Fontanet A, Chauty A. Findings in Patients From Benin With Osteomyelitis and Polymerase Chain Reaction-Confirmed *Mycobacterium ulcerans* Infection. *Clin Infect Dis*. July 21, 2014; epub ahead of print.

**Vincent QB**, Ardant M-F, Marsollier L, Chauty A, Alcaïs A. Defining and targeting high-risk populations in Buruli ulcer—Authors' reply. *Lancet Glob Health*. 2014 Nov;2(11):e630.

**Vincent QB** et al. L. Buruli ulcer of extreme severity maps to the beta-defensin locus on chromosome 8. To be submitted.

### Publications issues de collaborations au sein du laboratoire au cours de la thèse :

Sologuren I, Boisson-Dupuis S, **Vincent QB**, Pestano P, ...//..., Abel L, Casanova JL and Rodríguez-Gallego C. Partial recessive IFN- $\gamma$ 1 deficiency: genetic, immunological, and clinical features of fourteen patients from eleven kindreds. *Hum Mol Genet*. 2011; Apr 15;20(8):1509-23., : equal contribution.

Ouederni M, **Vincent QB**, Frange P, Scerra S, ...//..., Abel L, Casanova JL, Fischer A and Picard C, Major Histocompatibility Complex Class II expression deficiency caused by a founder RFXANK mutation – A survey of 35 patients. *Blood*. 2011 Nov 10;118(19):5108-18.

Moncada-Vélez M, Martinez-Barricarte R, ...//..., **Vincent QB**, ...//..., Casanova JL, Boisson-Dupuis S, Partial IFN- $\gamma$ 2 deficiency is due to protein misfolding and can be rescued by inhibitors of glycosylation. *Blood*, 2013 Oct 3;122(14):2390-2401.

Lanternier F, Pathan S, **Vincent QB**, ...//..., Lortholary O, Casanova JL, Picard C, Grimbacher B, Puel A. Deep dermatophytosis and inherited CARD9 deficiency. *NEJM*, 2013 Oct 31;369(18):1704-14

Gaschignard J, **Vincent QB**, Cobat A, Alcaïs A, Implicit hypotheses are hidden power droppers in family-based association studies of secondary outcomes. Submitted.

Belkadi A, Bolze A, Itan Y, **Vincent QB**, ...//..., Casanova JL, Abel L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Submitted.

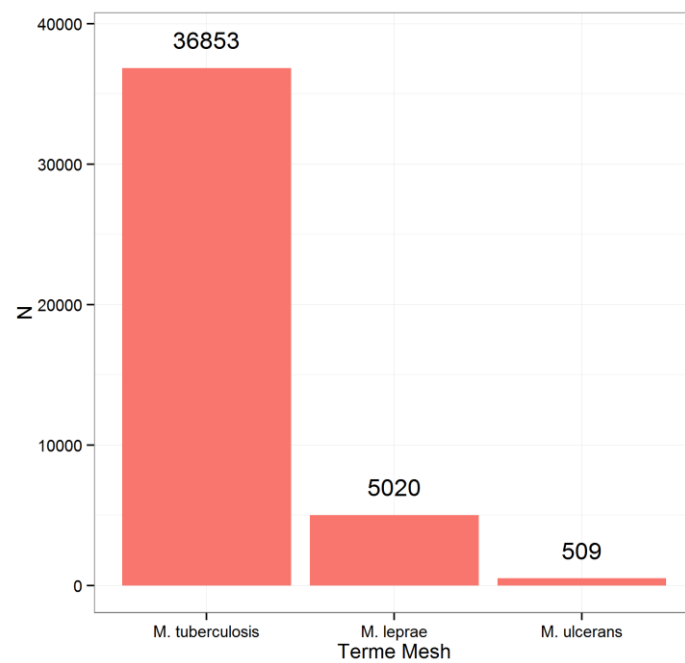


# Introduction :

## Etat des connaissances et motivations de l'étude

---





**Figure 1. Nombre d'articles référencés dans PubMed pour *M. tuberculosis*, *M. leprae* et *M. ulcerans***

**Recherche par index Mesh, 14 août 2014, [www.pubmed.com](http://www.pubmed.com)**

L'infection par *Mycobacterium ulcerans* (*M. ulcerans*), également nommée ulcère de Buruli (UB), est une maladie tropicale méconnue et négligée. D'un point de vue académique, il suffit pour s'en convaincre de constater que seules cinq cents publications sont référencées pour l'index Mesh "*Mycobacterium ulcerans*" dans la base de données internationale PubMed contre près de 37000 pour le terme Mesh "*Mycobacterium tuberculosis*" (Figure 1). Nous commencerons ce travail par une revue des connaissances, depuis une perspective historique jusqu'aux mécanismes physiopathologiques de l'infection, et passerons en revue l'épidémiologie, le mode de transmission et les moyens diagnostiques et thérapeutiques de l'ulcère de Buruli.

Après cet état des lieux, les motivations de notre étude seront détaillées. Cette thèse aborde successivement trois thèmes : (i) l'épidémiologie de l'ulcère de Buruli par l'étude de la cohorte de cas confirmés du Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB) de Pobè au Bénin ; (ii) la génétique mendélienne de l'ulcère de Buruli par l'étude d'une famille consanguine présentant une ségrégation mendélienne d'une forme extrême de la maladie et enfin, (iii) la génétique complexe de l'ulcère de Buruli par analyse d'association pangénomique.



## Etat des connaissances

### L'ulcère de Buruli : perspective historique

*" The right leg, from above the knee, became deformed with inflammation, and remained for a month in this unaccountable state, giving intense pain, which was relieved temporarily by a deep incision and copious discharge. (...) My strength was prostrated; the knee stiff and alarmingly bent, and walking was impracticable."*

*James Augustus Grant, A Walk Across Africa, 1861*

*"Leg began to swell 5 months ago. 2 months ago the upper ulcer began. 1 month ago the lower one began. (...) Extensive sloughing. Deep ulcer extending through subcutaneous tissue, fat & muscle, almost down to bone. Undermined."*

*Dr E.N. Cook, Mengo Hospital Notes, 1910*

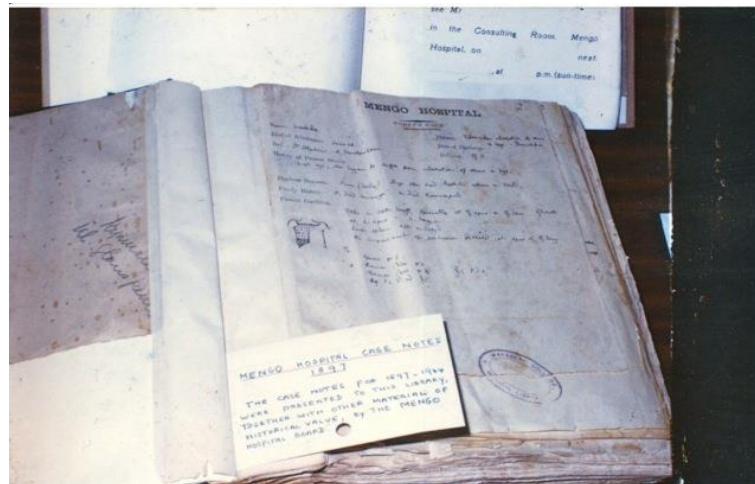
*"A new mycobacterial infection in man"*

*Dr P. MacCallum, Journal of Pathology and Bacteriology, 1948*

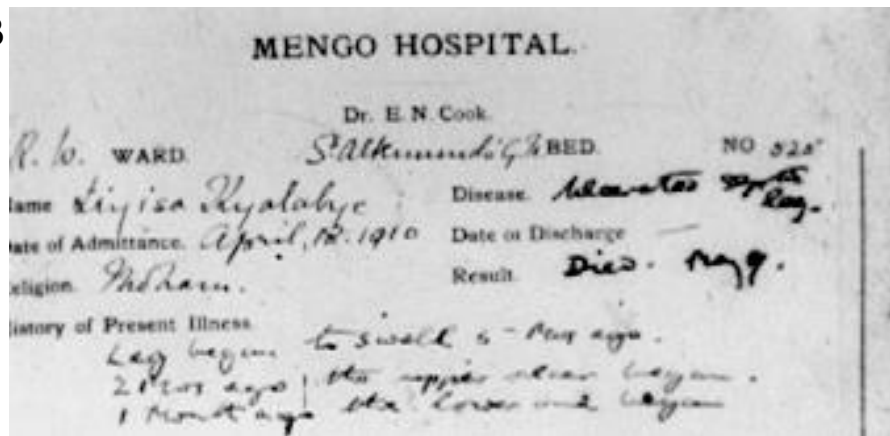
*"Twenty-eight of the forty patients so far recorded live in a very limited area within the Mengo district, known as Buruli."*

*Dr J.K. Clancey, The Lancet, 1961.*

A



B



C

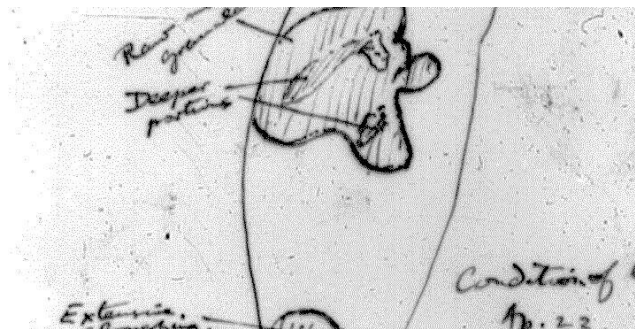


Figure 2. Descriptions cliniques originales de l'UB

(A) Mengo Hospital Case Notes (Dr A.R. Cook), 1897, ouvert à la deuxième page décrivant des ulcères tuberculeux des bras et jambes, considérés *a posteriori* comme UB

(B) Mengo Hospital Case Notes (Dr E.N. Cook), 1910, décrivant une forme œdémateuse secondairement ulcérée de la maladie

(C) Croquis annoté du cas décrit en B.

Photos gracieusement fournies par le Dr J. Hayman.

En 1861, l'explorateur écossais James Augustus Grant, à la recherche des sources inconnues du Nil en Afrique de l'Est, décrit la pathologie dont il est atteint. Elle est compatible avec une forme œdémateuse sévère de l'infection à *M. ulcerans* suivie d'une impotence fonctionnelle totale et fixée de la jambe. La première description médicale de l'UB est classiquement attribuée au docteur anglais Albert Ruskin Cook, qui à 26 ans débarque comme missionnaire anglican sur la côte kenyane. Il fonde en 1897 le premier hôpital de l'Ouganda, à Mengo, sa capitale, où il rédige des notes médicales qui constituent "l'une des merveilles médicales de l'Afrique" selon Davies (1). Il y décrit en particulier la maladie du sommeil, jusqu'alors inconnue en Afrique de l'Est (2).

Davies, qui a compilé ces notes, indique que les ulcères de jambe (sans plus de précision) sont responsables de 1 à 10% des admissions annuelles entre 1897 et 1956 (1). La deuxième note, datant du 28 juin (probablement 1897) rapporte le cas d'une patiente admise pour "ulcération tuberculeuse du bras et des jambes" ayant débuté 24 mois auparavant (Figure 2A). Son neveu, le Dr E.N. Cook rapporte en 1910 un cas sans équivoque : "La jambe a commencé à gonfler il y a 6 mois. Il y a 2 mois : apparition de l'ulcère supérieur. Il y a 1 mois: apparition de l'ulcère inférieur" (Figure 2B). Un croquis accompagne la description et est annoté : "Décharge majeure, ulcère profond, creusant à travers le tissu sous-cutané, la graisse et les muscles, presque jusqu'à l'os". Une légende pointe sur le bord de la lésion et commente : "décollé" (Figure 2C) (3).

En 1948, MacCallum, bactériologiste australien, décrit six cas d'ulcères cutanés à bords décollés, observés à Bairnsdale (300 km à l'est de Melbourne, Australie), et associés à une nouvelle espèce de bacille acido-alcool-résistant (BAAR), qui sera peu après nommé *M. ulcerans* (4). C'est la panne d'un incubateur qui a permis d'obtenir les premières cultures, dévoilant par ce détour du hasard une caractéristique fondamentale de la bactérie : une croissance quasi-exclusive à 32°C. Dès 1949, un cas est rapporté au Congo belge, dont la description clinique est en accord parfait avec un UB, et dont l'examen direct révèle de nombreux BAAR (5). En 1959, Janssens rapporte une série de 170 patients congolais présentant des ulcères chroniques de caractéristiques identiques à ceux

décrits par MacCallum et associée au même pathogène (6). En 1961, Clancey rapporte dans le Lancet 40 cas de ce qu'il nomme pour sa part "les ulcères de l'Ouganda" dont 28 proviennent d'une zone très restreinte du district de Mengo (là où Albert Cook décrivit les premiers cas 60 ans plus tôt), connue sous le nom de Buruli (7). Cette publication est à l'origine dès 1962 de la désignation actuelle de la maladie, ulcère de Buruli (UB), plus généralement nommée infection à *M. ulcerans* afin de prendre en compte les formes non ulcérées de la maladie (8). Un élément essentiel de la physiopathologie de *M. ulcerans*, la production d'une toxine cytotoxique, est suspecté dès 1965 suite à l'observation au plan anatomopathologique de l'extension de plages nécrotiques à distance des zones où la bactérie est identifiée (9). Elle ne sera purifiée qu'en 1999, et nommée mycolactone (10).

Par la suite, la maladie a été décrite dans plus de 30 pays en Afrique, Asie, Amérique Centrale et Amérique du Sud (Tableau 1 et Figure 3A). L'augmentation du nombre de cas observés et l'apparition de nouveaux foyers au cours des années 1980 et 1990 ont conduit l'Organisation Mondiale de la Santé (OMS) à lancer l'Initiative Mondiale pour l'Ulcère de Buruli (WHO Global Buruli Ulcer Initiative) en 1998. A l'issue de la première conférence de recherche et de contrôle de l'ulcère de Buruli (Conference on Buruli Ulcer Control and Research) organisée par l'OMS, les participants ont prononcé la Déclaration de Yamoussoukro (11), qui reconnaît le fardeau croissant que représente cette maladie émergente et négligée, en particulier en Afrique de l'Ouest, et souligne le faible niveau de connaissance sur cette pathologie. Les représentants des différents états se sont engagés à lutter contre la maladie. L'OMS a renouvelé son engagement contre l'UB en 2004 en adoptant une résolution en Assemblée Plénière (12).

**Tableau 1. Pays ayant déjà déclaré un cas d'UB**

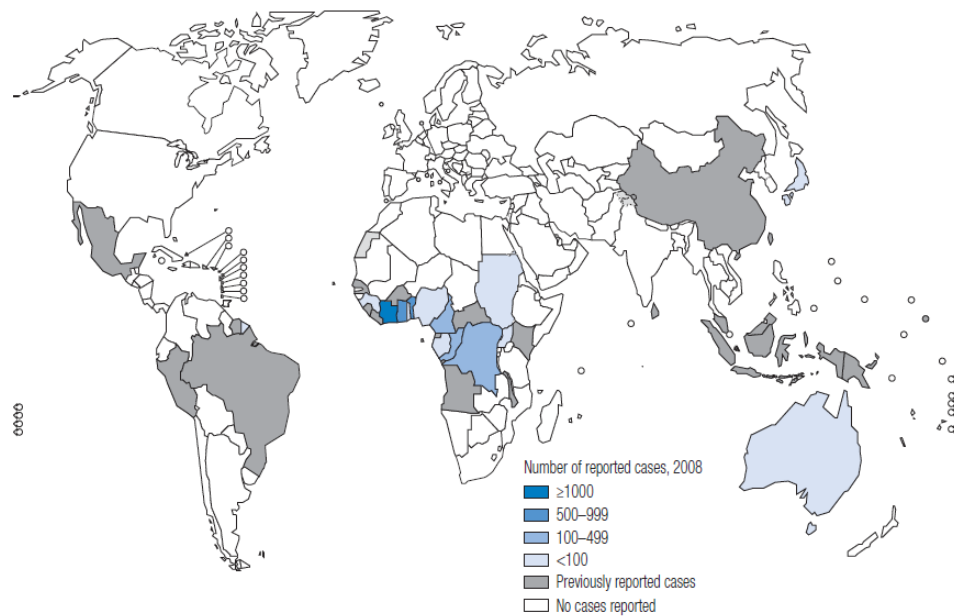
Continent	Pays
<b>Amérique</b>	Brésil, Guyane française, Mexique, Pérou, Suriname
<b>Afrique</b>	Angola, Bénin, Burkina Faso, Cameroun, Congo, Côte d'Ivoire, République centrafricaine, République Démocratique du Congo, Gabon, Ghana, Guinée, Guinée Equatoriale, Kenya, Libéria, Malawi, Nigéria, Ouganda, Sierra Leone, Soudan, Togo
<b>Asie-Pacifique</b>	Australie, Chine, Indonésie, Japon, Kiribati, Malaisie, Papouasie Nouvelle Guinée, Sri Lanka

## Epidémiologie de l'infection à *M. ulcerans*

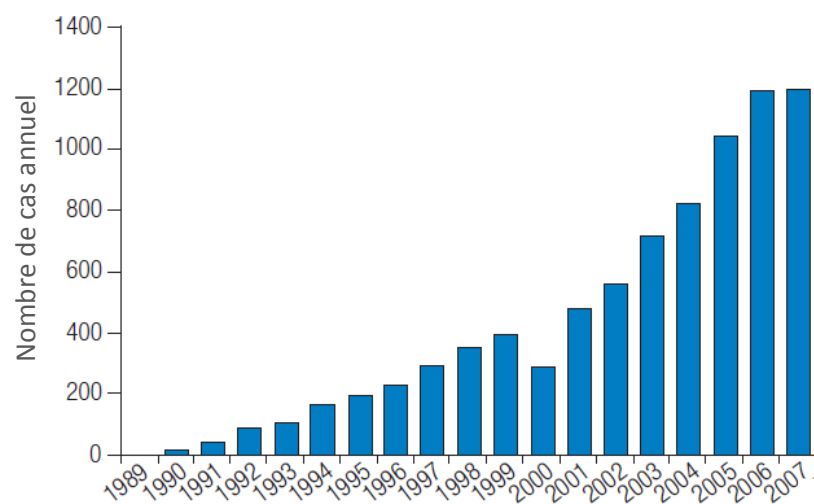
Bien que certains pays en zone tempérée soient marginalement affectés (Australie, en moyenne 60 nouveaux cas notifiés par an depuis 2002), l'infection à *M. ulcerans* sévit essentiellement dans les pays de la zone intertropicale, principalement d'Afrique subsaharienne (13). De fait, les trois pays les plus touchés actuellement sont la Côte d'Ivoire (en moyenne 1800 nouveaux cas annuels notifiés depuis 2002), le Ghana et le Bénin (en moyenne 900 nouveaux cas annuels notifiés par chaque pays depuis 2002) (Figure 3A, adaptée de (13,14)). Au Ghana, la prévalence nationale des lésions actives a été estimée à 20.7 pour 100 000 en 1999. Ce chiffre masque cependant des disparités majeures, comme dans tous les pays endémiques (15). Dans le district d'Amansie West, la prévalence atteint en effet 150.8 pour 100 000, tandis qu'une enquête révélait une prévalence de 22% dans le village de Tontokrom en 1993 (16). Récemment, une enquête au Ghana rapportait encore des prévalences de 15 à 20% dans plusieurs villages (17). Dans la région de Zou au Bénin, la prévalence était à 21.5 pour 100 000 (18). Récemment, une étude au Cameroun a confirmé une prévalence avoisinant 250 pour 100 000 dans le district d'Akonolinga (19). Au Ghana et dans la région de Zou au Bénin, la prévalence de l'UB dépasse donc celle de la lèpre (respectivement 9 pour 100000 et 13.4 pour 100000) pour atteindre celle de la tuberculose (respectivement 66 pour 100000 et 20.0 pour 100000) (15,18). Au plan mondial, l'UB est la troisième mycobactériose la plus fréquente. Le nombre annuel de cas notifiés à l'OMS est inférieur à 10000 mais sous-estime clairement le nombre de cas réels (13). La France n'est pas totalement épargnée, puisque la Guyane Française est le premier foyer d'Amérique du Sud, avec 195 cas de 1969 à 2003 (20).



**A**



**B**



**Figure 3. Epidémiologie de l'infection à *M. ulcerans***

(A) Distribution mondiale des cas rapportés en 2008 et des pays ayant déjà rapporté des cas par le passé

(B) Nombre annuel de cas rapportés au Bénin, 1989-2007.

Source : OMS (14)

La classification de l'UB comme maladie émergente repose sur l'identification seulement récente de cette pathologie, y compris dans un pays à haut niveau sanitaire comme l'Australie. Différents chiffres plus ou moins parcellaires suivant les pays montrent une augmentation d'incidence semblant dépasser un simple biais de notoriété. Au Bénin par exemple, le premier article indexé faisant mention de la maladie date de 1988 et 28 cas sont rapportés entre 1986 et 1988 (21,22). Sur l'ensemble du pays, la courbe de progression fournie par l'OMS entre 1989 et 2007 est éloquent (Figure 3B) (14). En Côte d'Ivoire, le pays actuellement le plus affecté par la maladie, aucun cas n'est décrit avant 1983 (23). Au Cameroun, la maladie a été décrite pour la première fois dans les années 1970 mais un nouveau foyer de la maladie s'est récemment développé dans la région de Bankim (24,25). Dans d'autres régions du monde, comme en Australie, mais dans des proportions plus modestes, la maladie progresse tant en nombre (cas sporadiques depuis les années 1930, 32 cas en 2002, 61 en 2007, 143 en 2011) qu'en territoire (confinée à Bairnsdale jusqu'en 1982, lieu de sa description originelle, puis apparition de plusieurs nouveaux foyers) (13,26). Il est intéressant de remarquer que les foyers historiques, ougandais en particulier (plus de 1000 cas en deux ans, 1969-1971), semblent moins actifs actuellement, avec moins de 100 cas déclarés en 2009 et plus aucun depuis (20,27,28). A l'échelle mondiale, une baisse du nombre de cas notifiés a été observée depuis 2008 (13). Les motifs de cette baisse sont inconnus, mais il faut garder à l'esprit que la qualité du recensement des cas varie fortement selon les pays et les périodes.

L'UB touche tous les âges de la vie mais en Afrique les études sur grands échantillons (plus de 1000 cas) estiment la médiane d'âge au diagnostic à 15 ans environ (29–31). Quatre études ont estimé les taux d'infection par tranche d'âge. Deux d'entre elles, au Bénin et au Cameroun, montrent une courbe bimodale (10-14 ans et plus de 50 ans), tandis que les deux autres, en Côte d'Ivoire et en Ouganda dans le camp de réfugié de Kinyara (suivi prospectif pendant 4 ans d'une population initialement totalement indemne déplacée dans une zone d'endémie), n'identifient qu'un seul pic, chez les 5-14 ans (23,27,30,32). Les effectifs sont néanmoins faibles au-dessus de 50 ans. La maladie touche les deux sexes de manière sensiblement équilibrée (15,29–31), même si plusieurs études au

Cameroun, au Ghana et en Ouganda rapportent une prédominance féminine parmi les cas adultes et masculine parmi les cas pédiatriques (16,27,31–35). L'épidémiologie de l'UB est sensiblement différente en Australie, où la médiane d'âge de 180 patients traités sur une période de 13 ans était de 61 ans (36).

Une dizaine d'études cas-témoin ont été menées au Ghana, au Bénin, au Cameroun, en Australie, en Côte d'Ivoire et en Ouganda afin d'identifier les facteurs de risque d'infection. Un contact physique régulier avec des sources d'eau naturelle à titre récréationnel ou domestique est un facteur de risque retrouvé régulièrement (33,34,37,38). Deux études en Côte d'Ivoire et en Australie n'ont pas identifié la natation comme facteur de risque, contrairement à deux études au Cameroun et au Ghana (23,34,38,39). L'observation que la proximité à l'eau (jusqu'à quelques kilomètres) est un facteur de risque d'UB, mais que le contact direct n'est pas une condition strictement nécessaire au développement de la maladie est rapportée par plusieurs auteurs (23,27). D'une manière générale, la source d'eau utilisée pour boire n'est pas associée à la maladie (16,33,40), même si une étude est discordante (41).

Le soin apporté au nettoyage des abrasions et plaies a été associé de manière répétée à une protection contre l'UB (33,34,39,40,42). Le port de vêtements longs (pantalons, chemise à manches longues, jupe couvrante) est également associé à une protection contre la maladie, avec des odds-ratios (ORs) conséquents (23,33,34,39). En Côte d'Ivoire, la protection par le port de pantalons longs était encore plus significative lorsque l'analyse était restreinte aux patients présentant des lésions au niveau des jambes (OR 0.15, IC95% [0.03-0.70]). Enfin, trois études sont en faveur d'une protection par les moustiquaires, et une en faveur d'une protection par l'utilisation de sprays répulsifs contre les insectes (34,39,40,42).

Jusqu'ici, l'infection à *M. ulcerans* ne semble pas être favorisée ou aggravée par l'infection par le virus de l'immunodéficience humaine (VIH). On trouve dans la littérature des cas VIH+ présentant une forme sévère avec atteinte osseuse disséminée (43–45) mais également des formes

d'évolution classique (46,47). Une étude cas-témoin au Bénin suggère une association VIH et UB, mais la proportion de VIH+ chez les témoins (0.3%, 2 individus sur 613) semble anormalement basse (48). Symétriquement, on trouve des cas VIH- d'évolution sévère (49–51) et une étude cas-témoin au Ghana ne retrouve pas d'association entre le VIH et l'UB (33). La fréquence des formes osseuses ne semble pas plus élevée chez les patients VIH+ que les patients VIH-, d'après l'étude d'une série de cas au Bénin (52). La question de l'association VIH-UB reste donc largement ouverte et est au centre des sujets mentionnés par une revue récente proposant des axes de recherche prioritaire dans l'UB (53). Ce sera l'un des thèmes de notre étude épidémiologique. La puissance des études publiées est modeste du fait du faible nombre de coinfections. L'UB touche en effet essentiellement des enfants en zone rurale alors que le VIH touche essentiellement les adultes des zones urbaines.

Au total, le tableau épidémiologie de l'UB peut se décrire relativement simplement : il s'agit d'une maladie émergente tropicale majoritairement pédiatrique, liée à la proximité de zones inondables ou fleuves à écoulement lent. Le port de vêtements couvrants, le nettoyage des plaies, et l'utilisation de moustiquaire ou répulsifs sont des facteurs protecteurs validés.



## L'infection à *M. ulcerans*, une maladie vectorielle ?

*" Est-ce qu'un sorcier voudrait tuer tous ces gens ?*

*Y compris les enfants, les tout petits, qui n'ont pas encore pu faire du mal ?"*

*Groupe de femme, Bénin (54).*

*"Dans mon cas je pense que c'est de la sorcellerie, parce que j'ai une bonne hygiène."*

*Ghana (55).*

Plus de cent ans après la description clinique de l'UB, plus de cinquante ans après la description microbiologique de *M. ulcerans*, sept ans après le séquençage de son génome, le mode de transmission de la maladie n'est pas clairement établi et la moitié des personnes interrogées dans les zones endémiques africaines attribue l'UB à de la sorcellerie.

Dans les pays ayant rapporté des cas d'UB, les foyers hyperendémiques sont situés en zone rurale de grande humidité : bordure de grands fleuves, lacs, marais, zones inondables (Figure 4A, adaptée de (56,57) ). Le rapport fondateur de Clancey note l'accumulation des cas sur les rives du lac Kyoga (7). L'étude de l'épidémie du camp de Kinyara installé sur les bords du Nil en 1964 (220 cas de 1965 à 1970, taux d'incidence cumulé de 8.7%) met en évidence un gradient croissant de cas à l'approche du fleuve (27). Les auteurs constatent l'extinction rapide de l'épidémie après le déplacement du camp à 240 kilomètres du Nil. L'observation d'une incidence maximale aux abords d'un cours d'eau a été faite plusieurs fois, en Afrique et ailleurs, en particulier par des approches géosatellitaires (19,23,34,58–60). Au Bénin, la prévalence de l'UB est corrélée à l'identification de sources d'eau contaminées par *M. ulcerans* (61).

**A**



**B**



**Figure 4. Transmission de *M. ulcerans***

(A) Zone d'hyperendémicité typique, Bénin

(B) Punaies d'eau *Naucoris flavicollis* (taille 1.5cm, à gauche) et *Belostoma cordofna* (taille 10 cm, à droite), vecteurs de *M. ulcerans*

Source: (56,57)

Plusieurs épidémies ont été attribuées à des conditions météorologiques exceptionnelles : inondations en Ouganda en 1964, inondations en Papouasie en 1951 (28). On a également rapporté l'émergence de foyers suite à la construction de retenues d'eau de dimensions diverses (par exemple, le petit lac d'agrément du campus de l'université d'Ibadan au Nigéria, 14 cas en 1972-1976) (62). Une analyse à l'échelle nationale en Côte d'Ivoire a également corrélié endémicité et barrages, en particulier barrages de petite taille, qui sont les plus fréquentés par la population (60). L'émergence du foyer de Bankim au Cameroun pourrait être liée à la construction d'un grand barrage à la fin des années 1980, mais ce lien est débattu (24,32,63). Une étude récente au Cameroun rapporte que *M. ulcerans* est fréquemment identifié dans les zones marécageuses apparaissant artificiellement le long des routes (63). Une épidémie en Australie (29 cas, 1993-1995) a été reliée à l'arrosage d'un golf par des eaux pompées dans un marais avoisinant dont le drainage avait été obstrué par des travaux de construction. Deux cas ont été rapportés dans l'année suivant l'arrêt du pompage, et plus aucun depuis (20,26).

La transmission interhumaine semble inexistante ou exceptionnelle. En dehors d'un chirurgien plastique en 1987, il n'y a aucune description de cas chez un soignant. De même, la maladie ne s'est pas maintenue après le déplacement des réfugiés de Kinyara vers un autre site, et ce malgré la présence de 24 nouveaux cas parmi les déplacés (27). Deux modes de transmission mutuellement non exclusifs ont été proposés dès 1961 et sont encore vivement débattus actuellement : une transmission directe depuis un réservoir environnemental inerte (eau, sol, surface de plantes) favorisée par un traumatisme mineur ou bien l'inoculation de la bactérie par un insecte vecteur. L'hypothèse d'un réservoir inerte de *M. ulcerans*, avec une transmission à l'homme à la faveur d'un traumatisme, est défendue par Meyers dès 1974 qui rapporte 14 patients avec un épisode traumatique clair au site de la première lésion (64). Plus récemment, cette hypothèse est étayée par la détection de l'ADN de *M. ulcerans* dans divers prélèvements environnementaux : eau, sol, boue (56). Néanmoins, une étude récente chez le cochon d'Inde rapporte que le dépôt de *M. ulcerans* sur une peau abrasée n'aboutit jamais au développement de lésions cutanées,

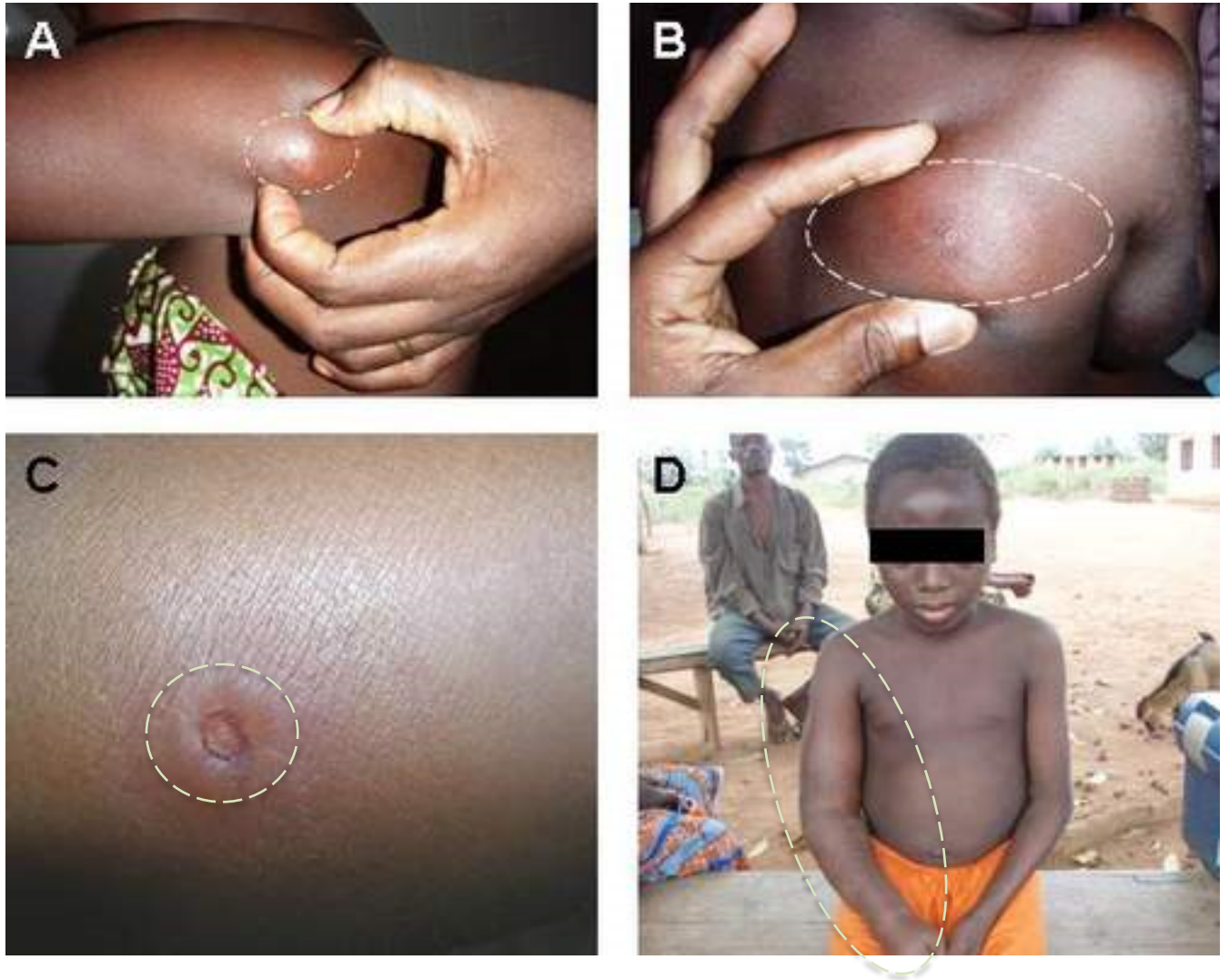


contrairement à une injection intra-dermique de la bactérie ou au dépôt de *Staphylococcus aureus* sur une peau abrasée (65). L'épidémie australienne autour du système d'arrosage du golf de Phillip Island près de Melbourne a également fait proposer une éventuelle transmission par aérosol (26). L'hypothèse d'une transmission par un insecte-vecteur (détaillée ci-dessous) a fait l'objet de très nombreuses investigations, et implique la punaise d'eau et peut-être le moustique.

Une série d'observations de terrain et d'expériences de laboratoire rendent probable l'implication de certaines punaises d'eau (essentiellement, de la famille des *Belostomatidae* et de la famille des *Naucoridae*, Figure 4B) dans la transmission de l'UB en Afrique sans pour autant estimer précisément la fraction de cas attribuables à ce mode de transmission. L'ADN de *M. ulcerans* a été identifié pour la première fois chez des insectes aquatiques en 1999 (66), et la bactérie a été cultivée également pour la première fois depuis une source environnementale en 2008, à partir d'un Hémiptère aquatique au Bénin (67). Depuis, *M. ulcerans* a été retrouvé dans de nombreuses familles d'insectes aquatiques en Afrique, au sein d'un cycle de transmission aquatique complexe (63,68–70). De manière convaincante, *M. ulcerans* a été identifié par réaction de polymérisation en chaîne (PCR) dans les glandes salivaires de punaise capturées en zones endémiques, mais pas en zones non-endémiques, dans une large étude au Cameroun (71). Au laboratoire, des punaises nourries par des repas contaminés sont colonisées par la bactérie, qui se multiplie par la suite dans les glandes salivaires. Ces punaises transmettent l'agent pathogène, puisqu'après morsure, la souris développe des lésions cliniques (72–74). Les punaises aquatiques peuvent voler et mordre l'homme, bien qu'elles ne soient pas hématophages. On les retrouve régulièrement à distance des points d'eau dans les villages (L. Marsollier, communication personnelle). Une étude sérologique chez l'homme a montré que dans les zones d'endémie, la population a développé une réponse humorale spécifique contre les protéines de glandes salivaires de punaise aquatique, objectivant les morsures par ces insectes, quand bien même elles ne soient qu'accidentelles (75).

L'hypothèse d'une transmission par les moustiques est relativement naturelle : ces insectes sont responsables de nombreuses pathologies infectieuses virales et parasitaires dans le monde (comme par exemple le paludisme, la dengue, la fièvre jaune, la filariose); la morsure est fréquente, reconnue par chacun, et non accidentelle puisque nécessaire à leur nutrition. En revanche, il est intéressant de remarquer que la transmission d'une bactérie par un moustique n'a pas été décrite à ce jour. Néanmoins, en Australie, l'hypothèse d'une transmission par le moustique est privilégiée. Une étude cas-témoin australienne a en effet observé que les patients se souvenaient plus fréquemment avoir été piqués par un moustique que les témoins et cette piqûre était localisée au site d'ulcération (39). L'ADN de *M. ulcerans* a été identifié dans une proportion faible mais non-nulle de moustiques capturés en zone d'endémie (4.3 moustiques pour mille) (76). Cette proportion est géographiquement corrélée à l'incidence de l'UB (77). Néanmoins, une étude en laboratoire ne parvient pas à infecter des moustiques adultes par *M. ulcerans*. De plus, si les PCR sur broyat de moustiques sont positives après contact avec des repas contaminés, ce n'est pas le cas des PCR sur glandes salivaires ou intestin, ce qui suggère une simple contamination externe de l'insecte (78).

En Afrique, aucune étude n'a rapporté la recherche de *M. ulcerans* chez les moustiques, tandis qu'en Australie, aucune étude n'a évalué la colonisation des insectes aquatiques. L'effet protecteur de la moustiquaire ne discrimine pas ces vecteurs. Un réservoir mammifère est également recherché. En Australie, les opossums sont un réservoir probable, puisque 41% des selles d'opossum collectées en zone d'endémie contiennent de l'ADN de *M. ulcerans*, contre <1% des selles collectées en zone non endémique (79). Il n'y a pas d'opossums en Afrique. Au total, il semble clair que la punaise d'eau est un vecteur de la maladie en Afrique. Il n'est en revanche pas prouvé que ce mode de contamination soit exclusif ou universel. L'hypothèse d'une transmission passive depuis l'environnement à la faveur d'une lésion cutanée reste d'actualité (32). Une étude en Afrique sur le point d'être publiée suggère qu'il y a multi-portages, que plusieurs routes de transmission sont possibles et que la transmission vectorielle est très minoritaire (JF. Guégan, communication personnelle).



**Figure 5. Formes pré-ulcératives de l'infection à *M. ulcerans***

(A) Nodule

(B) Plaque

(C) Papule en voie d'ulcération

(D) Œdème du bras droit

*Photos : Dr A. Chauty (80).*

## Aspects cliniques : diagnostic, évolution, séquelles

Le cas très particulier du camp de réfugié de Kinyara (implantation temporaire d'une population initialement indemne en zone d'endémie) et diverses observations ponctuelles ont permis d'estimer la durée moyenne d'incubation de la maladie entre deux et trois mois (27). Ainsi, alors que le camp a été ouvert fin 1964, les premières lésions sont apparues en février 1965. Par ailleurs, treize semaines après avoir quitté Kinyara pour le camp de Kyangwali, plus aucune lésion n'était déclarée. Une étude récente en Australie chez une vingtaine de patients vacanciers dont l'exposition est ponctuelle révèle des périodes d'incubation allant de 1 à 9 mois pour une moyenne de 4.5 mois (81).

Sur le plan clinique, la phase pré-ulcérate peut prendre quatre formes différentes (Figure 5): une papule érythémateuse non spécifique, observée essentiellement sur peau blanche en Australie; un nodule sous-cutané ferme et froid, de moins de trois centimètres de diamètre, solidaire de la peau mais libre des plans profonds; une plaque indurée, bien démarquée, surélevée, aux bords irréguliers, mesurant plus de trois centimètres de diamètre; une infiltration œdémateuse diffuse et ferme aux limites floues, ne prenant pas le godet, pouvant intéresser l'ensemble du membre (82). La coloration de la peau peut être modifiée au niveau des lésions. Ces lésions s'ulcèrent secondairement dans la majorité des cas. Une caractéristique sémiologique fondamentale de l'ulcération est son extension latérale profonde, laissant en surface de larges bords décollés, parfois nécrotiques. Le diamètre réel de la lésion, fréquemment au-delà de 15 cm, dépasse largement le diamètre apparent de l'ulcération et la lésion peut découvrir un membre ou un hémithorax entier (Figure 6). A l'inverse certaines lésions sont punctiformes. Le fond de l'ulcère, parfois suintant, peut avoir un aspect blanchâtre et cotonneux (Figure 6A-C). De manière surprenante, ces lésions sont indolores quelles que soient leur étendue, à l'exception des formes œdémateuses qui sont en général plus sévères et qui peuvent être douloureuses et fébriles. Les lésions touchent en majorité les membres, parfois le tronc, exceptionnellement le visage, et jamais les paumes et plantes de pied. Des adénomégalies peuvent être retrouvées. L'état général est conservé (57,83–86).



**Figure 6. Formes ulcérées, complications et séquelles**

- (A) Forme ulcéreuse située au niveau du coude. La périphérie de la lésion est hyperpigmentée témoignant d'une nécrose tissulaire en progression. En raison de la destruction du tissu adipeux, le fond de l'ulcère est jaune avec un aspect cotonneux.
- (B) Ulcération massive du coude. Les bords décollés témoignent de la progression de la lésion.
- (C) Forme disséminée avec atteintes osseuses. On notera la présence d'une forme œdémateuse au niveau de l'œil de ce jeune malade (flèche).
- (D) Surinfection à *Staphylococcus aureus* sur une forme ulcérée.
- (E) Cicatrice au niveau de la jambe après traitement chirurgical.
- (F) Séquelles invalidantes causées par des rétractions musculo-tendineuses (guérison spontanée)

Photos et commentaires : Dr A. Chauty (80).

Il est fréquent d'observer une dissémination métastatique de l'infection avec l'apparition d'une ou plusieurs autres localisations cutanées non contigües, ipsi- ou contra-latérales (6 à 30% des patients selon les séries, le plus souvent autour de 10%) (29,33,87). Dans 5 à 10% des cas, une ostéomyélite et parfois une arthrite à *M. ulcerans* se déclare (Figure 6C) (18). En l'absence de traitement, après une évolution de plusieurs mois à plusieurs années, les lésions cutanées régressent lentement, marquant la peau de larges cicatrices rétractiles (Figure 6E-F). Des rechutes sont possibles, sur la localisation initiale ou à distance, avec des taux extrêmement variables selon les études, de 6 à 47%. Il y a toutefois une ambiguïté dans certaines études entre rechute et absence de guérison initiale, ce qui peut expliquer les chiffres les plus élevés (29,88,89).

De manière remarquable, les surinfections restent rares malgré un environnement tropical et une hygiène rudimentaire (Figure 6D). Certaines données récentes suggèrent néanmoins que les surinfections pourraient être moins exceptionnelles qu'anticipé. Douze patients d'une série de 51 patients béninois hospitalisés ont reçu une antibiothérapie pour une suspicion clinique de surinfection. Il s'agit bien entendu d'une surestimation de l'incidence des surinfections puisque la majorité des patients ne sont pas hospitalisés. Une autre étude récente au Ghana rapporte 31 cas de surinfections documentées sur le plan bactériologique (essentiellement par *Staphylococcus aureus*, *Pseudomonas aeruginosa* et *Proteus mirabilis*) (90,91).



A



B



**Figure 7. Réactions paradoxales**

- (A) Une lésion ulcération initialement limitée (à gauche) ayant initialement bien répondu à l'antibiothérapie s'aggrave après 5 semaines de traitement. La lésion est chaude et suintante. Photos issues de (92).
- (B) Une lésion ulcération d'évolution lente depuis plusieurs mois chez un patient de 19 ans (à gauche) connaît une aggravation rapide et sévère 4 semaines après le début de l'antibiothérapie, accompagnée de signes généraux. Photos issues de (93).

En 2009, des cas d'aggravations secondaires des lésions après une amélioration initiale sous traitement ont été rapportés pour la première fois (94). Une lésion cicatrisée ou en voie de cicatrisation se ré-ouvre ou s'étend, généralement dans un délai de 3 à 12 semaines après l'initiation de l'antibiothérapie, mais parfois après plus d'un an (92,95,96). Dans certains cas, de nouvelles lésions se révèlent à distance de la lésion initiale. La sévérité du tableau clinique peut nettement dépasser celle de la présentation initiale (Figure 7). Historiquement, ces évolutions secondairement défavorables étaient considérées comme des rechutes liées à une mauvaise observance ou à un échec de l'antibiothérapie, mais la bonne réponse initiale à l'antibiothérapie et trois éléments atypiques de l'UB ont conduit à qualifier ces réactions de "paradoxaes" : cliniquement, la lésion est en général chaude, douloureuse et suintante; microbiologiquement, la culture reste négative; histologiquement, on retrouve un infiltrat inflammatoire marqué. La guérison nécessite un soin local de la plaie et pourrait bénéficier d'une corticothérapie systémique dans les cas les plus sévères, mais ne nécessite pas d'antibiothérapie (93,97). Ce phénomène, qui concerne entre 20 et 30% des patients d'après les deux premières études australiennes et ghanéennes, pourrait se rapprocher des réactions réverses de type I observées lors de la lèpre ou des syndromes de reconstitution immunitaire des coinfections VIH-tuberculose (92,94,95,98).

L'UB est exceptionnellement fatal, bien que quelques décès soient rapportés dans des circonstances très particulières (choc septique sur surinfection, tétanos, dénutrition sévère) (99,100). Ce sont les séquelles fonctionnelles et esthétiques majeures et permanentes qui font la gravité de cette maladie qui touche une population jeune (Figure 6E-F). La cicatrisation s'accompagne de la formation de brides et de rétractions tendineuses, d'une amyotrophie, d'un lymphoedème et d'une ankylose qui limitent l'amplitude articulaire (6). Différentes études rapportent des taux d'amputation de 2 à 10% avec un pic à 19% dans une série de 106 formes osseuses au Bénin (2008) (33,52,87,99,101). Deux études récentes au Ghana révèlent qu'environ 60% des patients ont une limitation d'amplitude articulaire à long terme (essentiellement coude, genou, poignet) (87,102). Toutes ces limitations ne se traduisent pas en incapacité dans les gestes de la vie quotidienne, ce qui



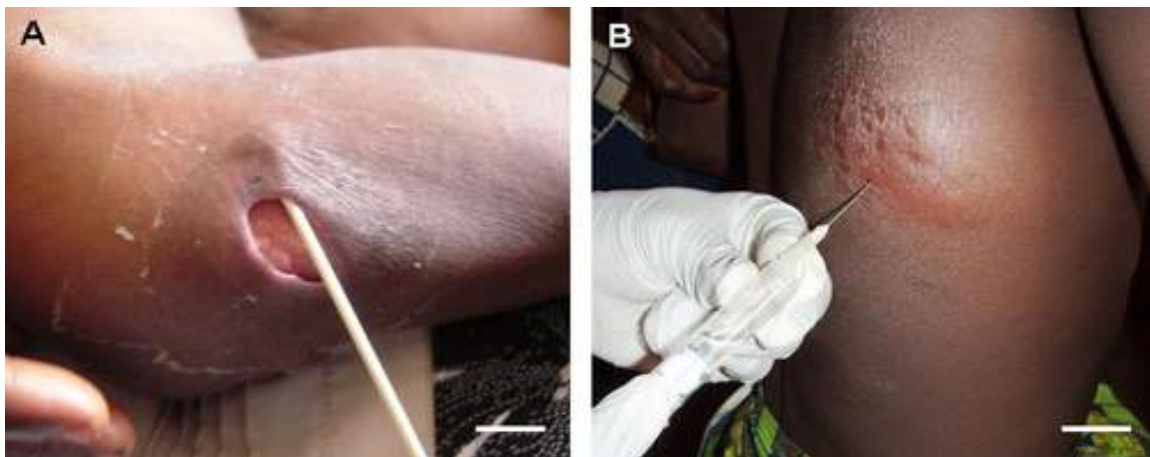
explique que d'autres études moins détaillées dans leur description du handicap rapportent des taux de séquelles plus bas compris entre 15 et 25%, mais correspondant à une incapacité évidente (23,34). Sur le plan fonctionnel, environ 30% des patients déclarent une impossibilité ou des difficultés pour aller chercher de l'eau au puits; 20% à verser de l'eau dans une tasse, 15% à se laver, 20% à marcher sur une surface plane, 40% à courir, 45% à transporter leur moisson (101).

Quels que soient les critères retenus, il est indiscutable que l'UB affecte significativement la trajectoire de vie de nombreux malades puisqu'environ 50 % des patients adultes doivent suspendre leur activité professionnelle (dont environ 80% sans retrouver de travail), et qu'environ 25% des enfants interrompent leur scolarité (23,87,102). Il existe en outre une certaine stigmatisation de la maladie, puisque dans une étude au Ghana, 40% des personnes interrogées n'accepteraient pas un ancien patient comme chef de village (103). Vingt pour cent des patients guéris, un tiers des patients en cours de traitement et la moitié des témoins interrogés considèrent que l'UB est un obstacle probable ou avéré au mariage (55,104). Ceci dit, cette stigmatisation n'est pas absolue puisque 90% des personnes interrogées accepteraient qu'un patient occupe la place d'instituteur. Les malades ne sont pas physiquement mis à l'écart ni globalement rejetés, comme ce fut et c'est encore le cas pour la lèpre (103).

## Diagnostic et prise en charge

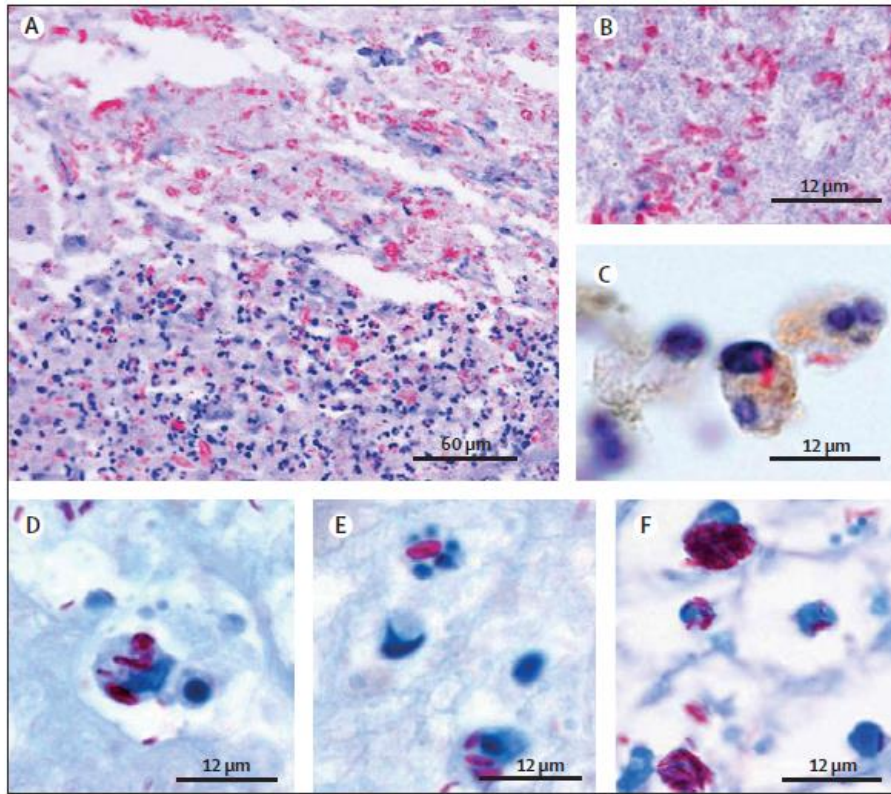
### Diagnostic positif et diagnostic différentiel

L'aspect de l'UB (ulcération indolore, bords décollés) et le contexte (zone d'endémie) rendent le diagnostic clinique des formes ulcérées relativement aisé. Les formes pré-ulcérales (nodules, plaques, oedèmes) sont de diagnostic moins évident. Les diagnostics différentiels incluent l'ulcère tropical phagédénique (infection polymicrobienne à fusobactéries et spirochètes à l'origine d'ulcères très douloureux d'évolution rapide), la tuberculose cutanée (dont l'aspect nodulaire ou ulcéré peut être trompeur, (105)), le pian (trépanomatose à l'origine d'ulcères bourgeonnants), la leishmaniose cutanée et certaines mycoses tropicales (82,106–108). En fonction du contexte, on pourra considérer des diagnostics différentiels non-infectieux, comme l'ulcère veineux, l'ulcère diabétique ou une pathologie carcinomateuse (82). Le diagnostic différentiel des formes oedémateuses inclut potentiellement l'érysipèle et la fasciite nécrosante qui évoluent néanmoins sur un mode aigu (82). L'OMS recommande la confirmation de l'infection à *M. ulcerans* par au moins une voire deux méthodes paracliniques, y compris dans les formes ulcérales (83). L'échantillon initial peut être un écouvillon réalisé en dessous des berges en profondeur, une biopsie per-chirurgicale ou une aspiration à l'aiguille fine d'un nodule ou d'une plaque (Figure 8).



**Figure 8. Confirmation diagnostique - prélèvements**

- (A) Prélèvement par écouvillonnage sur une forme ulcérée. Le prélèvement doit être réalisé en profondeur sous les bords décollés. Il est conseillé de réaliser plusieurs prélèvements à différents endroits.
- (B) Prélèvement à l'aiguille fine. Cette méthode de prélèvement peu invasive et peu douloureuse est recommandée pour les lésions fermées. Photos et commentaires : Dr A. Chauty (80).



**Figure 9. Aspects anatomopathologiques de l'infection à *M. ulcerans***

Biopsie d'un patient africain

(A) Les BAAR sont principalement trouvés dans la région supérieure nécrotique, tandis que la région inférieure montre un infiltrat inflammatoire.

(B) Un détail de la région nécrotique montre des BAAR extracellulaires.

(C) Bacille intramacrophagique de la région inflammatoire du panneau A.

(D-E) Bacilles intracellulaires, provoquant l'apoptose (E, fragmentation du noyau)

(F) Regroupement massif de bacilles intracellulaires

Source : (109)

L'examen direct d'un frottis après coloration de Ziehl-Neelsen est un moyen rapide et peu coûteux de mettre en évidence les BAAR, mais il est peu sensible (40%) et non spécifique d'une mycobactérie donnée (57). La mise en culture sur milieu de Löwenstein-Jensen à 32°C est possible, mais longue (six semaines) et difficile (contaminations), avec une sensibilité d'environ 20-60%, similaire à l'examen direct, mais en revanche une spécificité totale pour *M. ulcerans*.

La méthode la plus sensible (98%) et spécifique (100% en l'absence de contamination du matériel par des échantillons antérieurs), mais souvent difficile d'accès dans les zones d'endémie, est la PCR amplifiant la séquence d'insertion *IS2404* (57,83). Il faut cependant remarquer que cette méthode ne permet pas un suivi du traitement, puisqu'elle reste positive par amplification de débris microbiens, bien après la guérison clinique.

Enfin, l'examen anatomopathologique d'une biopsie peut révéler des aspects caractéristiques de l'infection à *M. ulcerans* avec une sensibilité élevée de l'ordre de 90%. On retrouve une nécrose massive du tissu graisseux sous-cutané (avec cellules fantômes - adipocytes nécrosés), un grand nombre de BAAR essentiellement extracellulaires, un infiltrat inflammatoire faible en comparaison de la charge bacillaire, restreint aux frontières de la lésion, et une atteinte vasculaire inflammatoire thrombotique (Figure 9, adaptée de (109)).

Par ailleurs, le diagnostic de réaction paradoxale repose sur (i) l'évolution secondairement défavorable d'une lésion d'UB traitée par antibiotique, ou à l'apparition d'une nouvelle lésion à distance du site initial et (ii) un examen anatomopathologique identifiant un infiltrat inflammatoire intense (82,92).

## Prise en charge

Le traitement de l'UB a connu un bouleversement complet ces dix dernières années suite à différents rapports démontrant l'efficacité d'une antibiothérapie parfois exclusive dans le traitement de la maladie. Jusqu'au début des années 2000, l'opinion dominante était que seule une exérèse large des lésions avec marges saines pouvaient guérir le patient et prévenir les rechutes ou les disséminations (110). En effet, un premier essai thérapeutique en 1973 sur 106 patients n'a pas montré d'efficacité sur le nombre d'opérations chirurgicales requises ni sur la durée d'évolution d'un traitement par clofazimine en monothérapie (111). Le cotrimoxazole, puis l'association rifampicine-dapsone n'ont pas non plus montré d'efficacité (112,113). Pourtant, *in vitro*, *M. ulcerans* était sensible à la rifampicine, aux aminosides, aux macrolides, et aux quinolones (57). Par ailleurs, plusieurs publications ont également suggéré l'efficacité d'une antibiothérapie chez la souris, ce qui a relancé la question de l'antibiothérapie chez l'homme.

En 2005, un premier groupe a montré que l'administration chez l'homme de rifampicine (10 mg/kg/j) et streptomycine (15 mg/kg/j) pendant quatre semaines avant l'exérèse d'une plaque ou d'un nodule non ulcéré inhibait toute culture bactérienne sur le tissu d'exérèse, brisant le dogme selon lequel l'étendue de la nécrose rendait *M. ulcerans* inaccessible aux antibiotiques (114). Dans cette étude de petite taille (21 patients), la majorité des lésions avait régressé avant la chirurgie, et aucune n'avait progressé. Au Bénin en 2007, 219 patients présentant une lésion évocatrice ont reçu exclusivement ce traitement antibiotique pendant quatre semaines (100). A quatre semaines, l'état clinique de tous les patients s'était amélioré, de façon si spectaculaire pour 40% (n=91) d'entre eux qu'aucune chirurgie n'a été jugée nécessaire, alors que treize patients avaient initialement présenté des ulcères de plus de 15 cm. L'antibiothérapie exclusive a été poursuivie quatre semaines supplémentaires chez 122 patients (91 + 31 ayant refusé la chirurgie). Au total 102 patients (47 %) n'ont pas requis d'intervention chirurgicale. Environ 80% des ulcères de moins de 5 cm ont été traités par antibiothérapie seule, contre 20 % des ulcères de plus de 15 cm. Un an après la fin du traitement,

trois rechutes seulement étaient comptabilisées (dont deux chez des patients ayant reçu une antibiothérapie exclusive), soit une proportion inférieure aux taux de rechutes rapportés précédemment dans le cadre d'une prise en charge par chirurgie exclusive (88,111).

L'injection intramusculaire quotidienne de streptomycine expose à deux problèmes principaux : l'asepsie du site d'injection en zone rurale tropicale reculée, et la toxicité cochléo-vestibulaire. Plusieurs études ont très récemment cherché à réduire ou remplacer l'usage de la streptomycine dans le traitement de l'UB. En 2010, un essai thérapeutique au Ghana comparait chez 150 patients au total l'efficacité sur les lésions de moins de 10 cm de l'association de la rifampicine pendant huit semaines avec soit la streptomycine pendant huit semaines, soit la streptomycine pendant quatre semaines suivie de la clarithromycine pendant quatre semaines (115). Il a conclu à l'équivalence des deux procédures, résultat intéressant car diminuant le nombre d'injections nécessaires au traitement. Cet essai objective par ailleurs une perte auditive significative à long-terme chez 30% des patients (116). Enfin, en 2011, une série de 30 patients présentant des lésions de moins de 10 cm a été traité avec succès par un traitement oral exclusif par rifampicine et clarithromycine pendant huit semaines (117). Cinquante pour cent de ces patients n'ont pas eu recours à la chirurgie, 37 % à une chirurgie limitée (curetage, exérèse) et 13 % à une chirurgie lourde (exérèse, greffe de peau). En Australie, une série de 30 patients a également été traitée avec succès par l'association rifampicine-ciprofloxacine (118). Un essai thérapeutique est en cours au Ghana et au Bénin pour évaluer formellement l'efficacité du traitement oral par clarithromycine et rifampicine dans l'UB (référéncé sur [clinicaltrials.gov](https://clinicaltrials.gov), NCT01659437).

TABLEAU 3. CATÉGORIES ET BUTS DU TRAITEMENT, NIVEAU DU SYSTÈME DE SOINS ET DIAGNOSTIC REQUIS

Catégorie traitement	Forme de la maladie	Traitement	But premier	But secondaire	Niveau du système de soins	Diagnostic
Catégorie I :	Petite lésion unique (ex : nodule, papule, plaque et ulcère < 5 cm de diamètre)	Terminer les antibiotiques  Sur une articulation ou à proximité, maintenir la même mobilité que sur le côté indemne  N'envisager une intervention chirurgicale sur un site non critique qu'après 8 semaines de traitement antibiotique	Guérison sans chirurgie  Guérison sans restriction des mouvements	Rechutes évitées ou réduites	Communauté, centres de santé, hôpitaux de district	Diagnostic clinique solide (avec ou sans confirmation du laboratoire)
Catégorie II :	Plaque et formes œdémateuses non-ulcérées ou ulcérées  Lésion ulcérée étendue unique, 5–15 cm de diamètre	Terminer les antibiotiques avant une intervention chirurgicale, si possible  Sur une articulation ou à proximité même mobilité que sur le côté indemne	Guérison sans chirurgie  Réduction de l'étendue du débridement chirurgical éventuellement nécessaire  Guérison sans restriction des mouvements	Rechutes évitées ou réduites	Centres de santé, hôpitaux de district ou de niveau tertiaire	Diagnostic clinique solide (avec ou sans confirmation du laboratoire)
Catégorie III :	Lésions sur la tête et sur le cou, notamment au visage  Formes diffuses et mixtes, comme l'ostéite, l'ostéomyélite, une atteinte articulaire  Lésions multiples et ostéomyélite  Lésion étendue > 15 cm de diamètre	Terminer les antibiotiques avant une intervention chirurgicale, si possible  Sur une articulation ou à proximité maintenir la même mobilité que sur le côté indemne	Guérison sans chirurgie  Guérison sans restriction des mouvements	Rechutes évitées ou réduites	Hôpitaux de district et de niveau tertiaire	Diagnostic clinique solide (avec ou sans confirmation du laboratoire)

Tableau 2. Recommandations OMS pour la prise en charge de l'UB

Source : OMS (82)

Les recommandations officielles de l'OMS pour la prise en charge de l'UB sont résumées dans le Tableau 2 et dépendent partiellement de la présentation clinique (82). Toutes les formes cliniques bénéficient de huit semaines d'antibiothérapie. Le traitement le plus consensuel reste l'association streptomycine et rifampicine. Dans l'attente des résultats de l'essai clinique en cours, l'association rifampicine et clarithromycine (ou rifampicine et ciprofloxacine) est principalement recommandée chez la femme enceinte. Toutes les formes ulcérées bénéficient de soins locaux, par lavage au savon doux, détersion mécanique et pansement humide. La détersion nécessite parfois un curetage chirurgical simple. En fonction de l'évolution sous traitement antibiotique et de la taille des lésions, une chirurgie réparatrice avec greffe de peau est indiquée. Dans la mesure du possible, la chirurgie sera programmée après 4 à 8 semaines d'antibiothérapie. Les cas d'ostéomyélites peuvent nécessiter des chirurgies plus invasives. Une rééducation fonctionnelle est toujours nécessaire pour prévenir ou lutter contre les limitations d'amplitude articulaire (82). En cas de réaction paradoxale, l'antibiothérapie n'est pas modifiée. Les soins locaux et potentiellement chirurgicaux sont poursuivis ou repris. Dans les formes les plus sévères, une corticothérapie systémique avec décroissance progressive sur 4 à 8 semaines peut être prescrite.

Il n'existe pas actuellement de vaccin efficace à long terme contre l'UB (119). La vaccination par le BCG a montré une certaine efficacité chez la souris et chez l'homme, mais cette efficacité ne persiste pas au cours du temps. A leur arrivée en zone d'endémie à Kinyara en 1967, les personnes non vaccinées par le BCG ayant une intra-dermoréaction négative à la tuberculine ont été randomisées pour être vaccinées ou non vaccinées (120). Sur 16 mois de suivi, 3.4% des 606 vaccinés développaient un UB contre 7.1% des 624 non-vaccinés ( $p=0.007$ ). A six mois de suivi, la protection était plus importante avec 1.1 % de cas chez les vaccinés contre 4.2% chez les non-vaccinés. En 1976, un deuxième essai en Ouganda sur 8000 personnes avec un suivi de quatre ans est arrivé aux mêmes conclusions (121). Récemment, une étude cas-témoin rétrospective au Bénin s'est basée sur la présence d'une cicatrice de la vaccination par le BCG dans une population en théorie systématiquement vaccinée à la naissance, et montre que la réponse au BCG (telle que mesurée par



la présence d'une cicatrice) ne semble pas avoir d'effet protecteur (122). Dans une série de 372 cas d'UB, 11% des 304 patients présentant une cicatrice du BCG ont développé une forme osseuse contre 31% de ceux n'en présentant pas (123). Les auteurs concluent que le BCG protège plus particulièrement contre les formes osseuses de l'infection à *M. ulcerans*.

L'univers cartésien des essais comparatifs et de la médecine fondée sur les preuves est bien éloigné du quotidien des patients de ces zones d'Afrique rurale, où, à titre d'exemple, l'eau courante est rare et l'électricité absente ou intermittente. Les trois quarts des patients considèrent que le premier traitement de l'UB est l'application d'herbes prescrites par un médecin traditionnel (103). Le caractère chronique de cette plaie est un signe d'ensorcellement pour bon nombre d'entre eux, qui, dès lors, considèrent que le traitement ne relève pas de la médecine occidentale et font appel à l'oracle ou au sorcier. Ceux-ci prescriront éventuellement des restrictions comportementales (régimes, tabous sexuels) (54,55). Par ailleurs, la perspective de l'amputation effraie notablement les malades et les pousse à retarder leur consultation, y compris pour des lésions minimales (54,55,104).

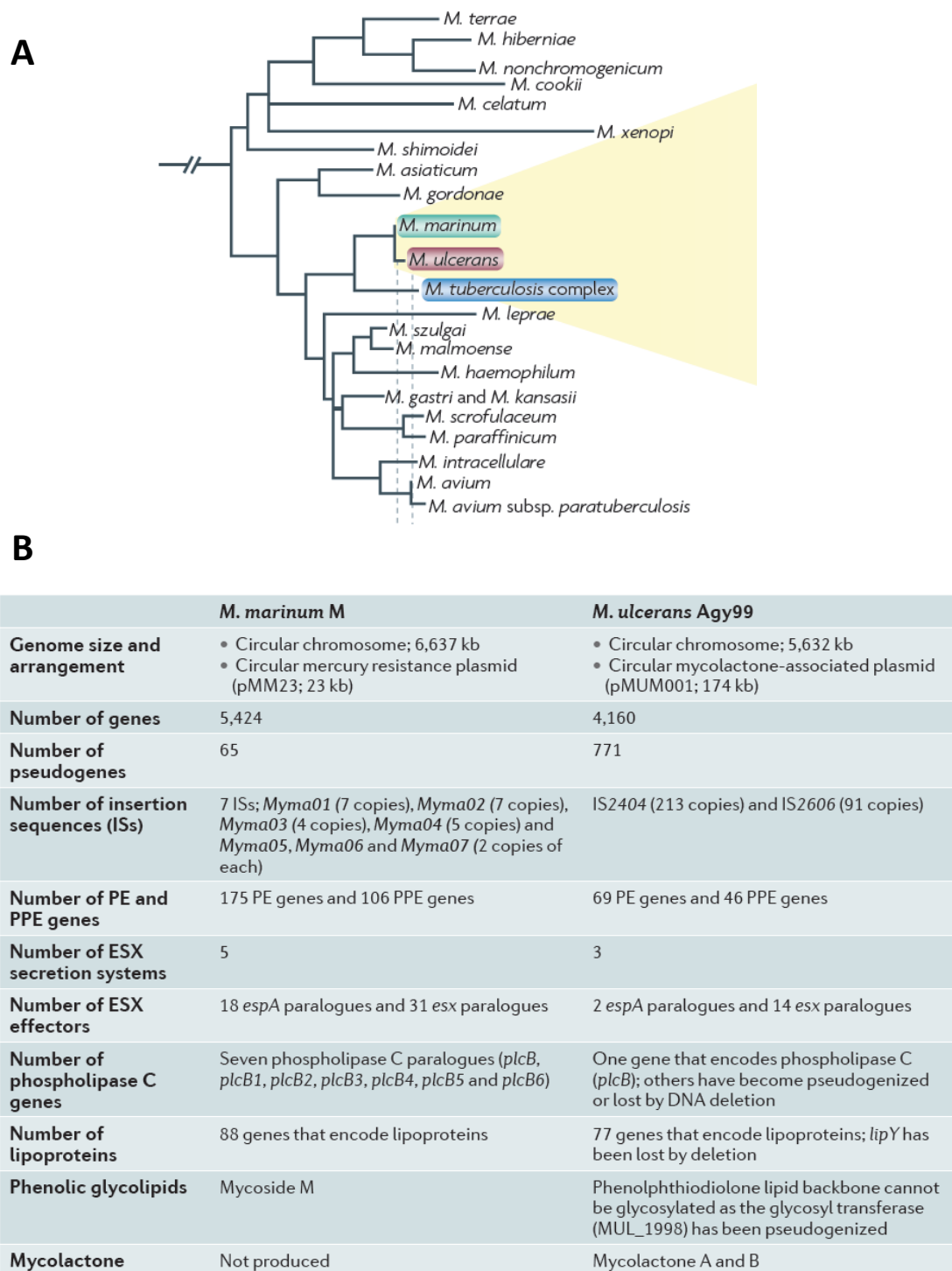
Enfin, le coût du traitement est conséquent, dans un contexte de protection sociale inexistante. Il faut ajouter au coût direct le coût indirect d'une perte de revenu lié à la présence de l'accompagnant. Un enfant hospitalisé est accompagné d'un membre de la famille qui devra le prendre en charge entièrement pendant la durée de l'hospitalisation (achat et préparation des repas, blanchisserie, toilette). En 1996, au Ghana, le coût moyen du traitement représentait 139 fois le revenu moyen journalier d'un travailleur agricole, tandis qu'en 2008 au Cameroun, le traitement de l'UB absorbait 25% du revenu annuel du foyer (99,124). L'avènement récent de traitements oraux ambulatoires exclusifs pour les formes modérées pourrait modifier notablement le délai de consultation des patients et limiter le fardeau que représente pour eux la prise en charge et les séquelles de l'UB.

## Physiopathologie de l'infection à *M. ulcerans*

*M. ulcerans* est apparenté à *M. marinum*, dont il diffère néanmoins significativement par différents aspects (125). Sur le plan clinique, *M. marinum* est à l'origine de rares infections cutanées localisées de type granulome des piscines (126). Sur le plan microbiologique, *M. ulcerans* a un temps de doublement de 50 heures contre 4 à 10 heures pour *M. marinum*, et ne produit pas de pigments photochromogènes (127,128). La composition de la paroi mycobactérienne semble significativement modifiée chez *M. ulcerans* (125). *M. marinum* a un mode de vie intracellulaire, tandis que *M. ulcerans* semble avoir un mode de vie mixte intracellulaire et extracellulaire. *M. ulcerans* présente la caractéristique unique parmi les mycobactéries de produire une toxine hydrophobe, la mycolactone.

### L'analyse du génome de *M. ulcerans* révèle une évolution réductrice

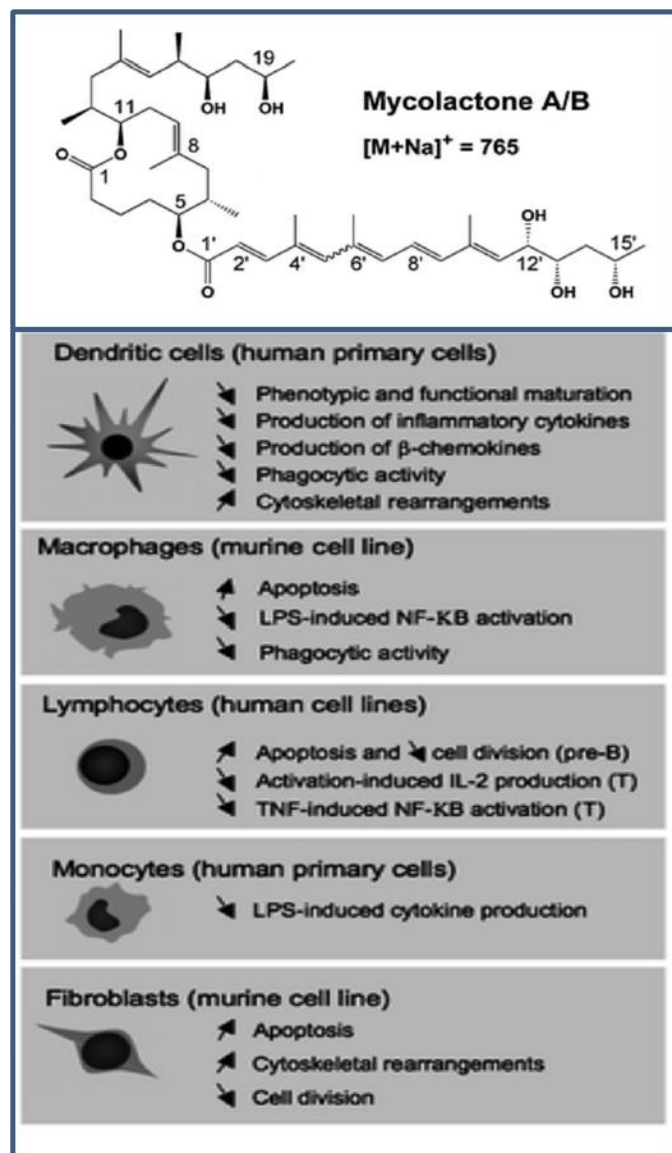
L'analyse du génome de *M. ulcerans* et *M. marinum* a révélé qu'ils descendent d'un ancêtre commun, avec une identité de séquence de 98.3%, à comparer à une identité de séquence de 78.5% entre *M. ulcerans* et *M. tuberculosis* (Figure 10A). Ils ont un total de 4000 gènes en commun. Néanmoins, le génome de *M. ulcerans* a subi une évolution dite "réductrice" (Figure 10B) : en comparaison avec *M. marinum*, *M. ulcerans* a perdu plus d'1.1 mégabases (Mb) de séquence (sur 6.5 Mb) et plus de 1000 gènes. L'insertion d'environ 200 séquences *IS2404* (Insertion Sequence 2404, utilisée pour la PCR diagnostique) a transformé une centaine de gènes en pseudogènes. Certaines protéines de surface de la famille PE/PPE (nommées d'après leur motif conservé Proline (P) - Glutamine (E) et Proline-Proline-Glutamine), certaines protéines du système d'excrétion ESX, responsables de l'export de l'antigène ESAT-6 (Early Secreted Antigenic Target 6), et certains composants hydrophobes de la paroi, dont les glycolipides phénoliques, ont disparu. Ces molécules sont pourtant considérées comme des facteurs de virulence majeurs des mycobactéries. Par ailleurs, suite à des délétions enzymatiques, *M. ulcerans* est incapable de produire des caroténoïdes, en faveur d'un mode de vie dans un environnement protégé de la lumière.



**Figure 10. *M. ulcerans* est proche de *M. marinum* mais a subi une évolution réductrice**

- (A) Arbre phylogénétique des mycobactéries (ARNr 16S).
- (B) Comparaison point par point de *M. ulcerans* et *M. marinum*.

Adapté de (128)



**Figure 11. La mycolactone, toxine diffusible cytotoxique et immunomodulatrice de *M. ulcerans***

(A) Structure chimique de la mycolactone A/B, avec son noyau lactone et ses deux chaînes latérales

(B) Résumé des activités cytotoxiques et immunomodulatrices de la mycolactone *in vitro*

*Adapté de (129)*

Ces pertes révèlent l'adaptation à une niche écologique particulière, probablement rendue possible par l'acquisition d'un plasmide géant de 174 kilobases (kb), pMUM, codant pour plusieurs enzymes nécessaires à la production de mycolactone (130). Différentes souches de *M. ulcerans* produisent différentes mycolactones (A à F), la plus active étant la mycolactone A/B, produite par les souches africaines (Figure 11). Elles ont toutes en commun un noyau cyclique lactone, classant la mycolactone dans la famille des macrolides, et deux chaînes latérales de type polycétide (129).

### La mycolactone : facteur de virulence majeur de *M. ulcerans*

L'histologie typique de l'infection à *M. ulcerans* retrouve de larges plages de nécrose de la graisse du tissu sous-cutané, dans lesquelles prolifèrent des quantités impressionnantes de bacilles majoritairement extracellulaires, regroupés en grappes par la production d'une matrice extracellulaire, contrastant avec un infiltrat inflammatoire périphérique de faible abondance (Figure 9) (131). On observe également une vasculopathie avec endartérite et thromboses. La mycolactone est probablement la principale cause de ces lésions. En effet, chez le cochon d'Inde, l'injection sous-cutanée de mycolactone purifiée reproduit les lésions provoquées par l'inoculation sous-cutanée de *M. ulcerans*, tant sur le plan clinique que microscopique, tandis qu'une souche mutée de *M. ulcerans* ne produisant pas la toxine ne provoque aucune lésion clinique (10).

La mycolactone sécrétée par *M. ulcerans* diffuse dans les lésions et peut être identifiée dans le sang de patients infectés (132,133). A l'échelle cellulaire, elle diffuse passivement à travers les membranes dans le cytoplasme et montre *in vitro* (kératinocytes, fibroblastes, adipocytes, lymphocytes, monocytes, macrophages, cellules dendritiques, neutrophiles) et *in vivo* (souris, cochon d'Inde, homme) des propriétés hautement cytotoxiques et pro-apoptotiques (134). Dans les cellules épithéliales, cette cytotoxicité est médiée par la liaison de la mycolactone à la protéine N-WASP qui enclenche une réorganisation majeure du cytosquelette, aboutissant à une perte d'adhérence

cellulaire et au déclenchement d'un programme de mort cellulaire (135). Dans les cellules immunitaires, non-adhérentes, le mécanisme n'est pas précisément connu mais pourrait être médié par une protéine proche, WASP, dont l'activation provoque une instabilité génomique et une apoptose dans le contexte de certains déficits immunitaires rares de l'enfant (135).

En plus de sa cytotoxicité, la mycolactone possède à concentration plus faible, *in vitro*, un rôle immunomodulateur très polymorphe (Figure 11, adaptée de (129)), affectant monocytes, macrophages, cellules dendritiques et lymphocytes T. La mycolactone peut inhiber la production de la cytokine proinflammatoire TNF $\alpha$  (Tumor Necrosis Factor) par les monocytes et les macrophages et moduler la production de multiples cytokines par les monocytes et les cellules dendritiques (136–138). L'efficacité de la phagocytose par les macrophages et les cellules dendritiques est altérée par la mycolactone, ainsi que le transfert des cellules dendritiques et des lymphocytes T naïfs vers les ganglions, et donc l'activation de la réponse immunitaire adaptative (129,136,138–141). La pertinence biologique *in vivo* de ces effets reste incertaine car la concentration réelle de mycolactone dans les lésions est inconnue. L'infection par *M. ulcerans* déclenche une réponse immunitaire qui permet le contrôle de l'infection ceci même en l'absence de traitement. Notamment, alors qu'on attribue à la mycolactone un effet immunosuppresseur (142), il n'y a pas d'infections opportunistes dans l'histoire naturelle de l'UB. Chez la souris, il a été démontré que l'effet immunosuppresseur est en réalité local (au niveau de la lésion) et régional (au niveau des ganglions lymphatiques) mais pas systémique (143,144).

La mycolactone est également impliquée dans la physiologie d'une des caractéristiques les plus étonnantes de l'UB. Les lésions provoquées par *M. ulcerans* sont indolores malgré une extension parfois spectaculaire. Cette caractéristique est à l'origine de retards de consultation et de prise en charge importants. Historiquement, elle était attribuée à une destruction neuronale par la mycolactone dans le tissu lésé. Très récemment, il a été démontré chez la souris que cette destruction neuronale n'existait pas, mais que la mycolactone provoquait une hyperpolarisation de la

membrane des neurones induisant une analgésie. Cette hyperpolarisation fait suite à l'activation du récepteur 2 de l'angiotensine, AT2R, par la mycolactone. L'activation de la voie de l'acide arachidonique par AT2R aboutit à un efflux potassique hyperpolarisant via les canaux potassiques de type TRAAK (145). Quinze ans après l'identification de la mycolactone, deux cibles moléculaires ont donc été identifiées, les protéines WASP/N-WASP impliquées dans la mort cellulaire programmée des tissus épithéliaux et AT2R responsable de l'analgésie observée dans l'UB.

### *M. ulcerans* : un germe intracellulaire ou extracellulaire ?

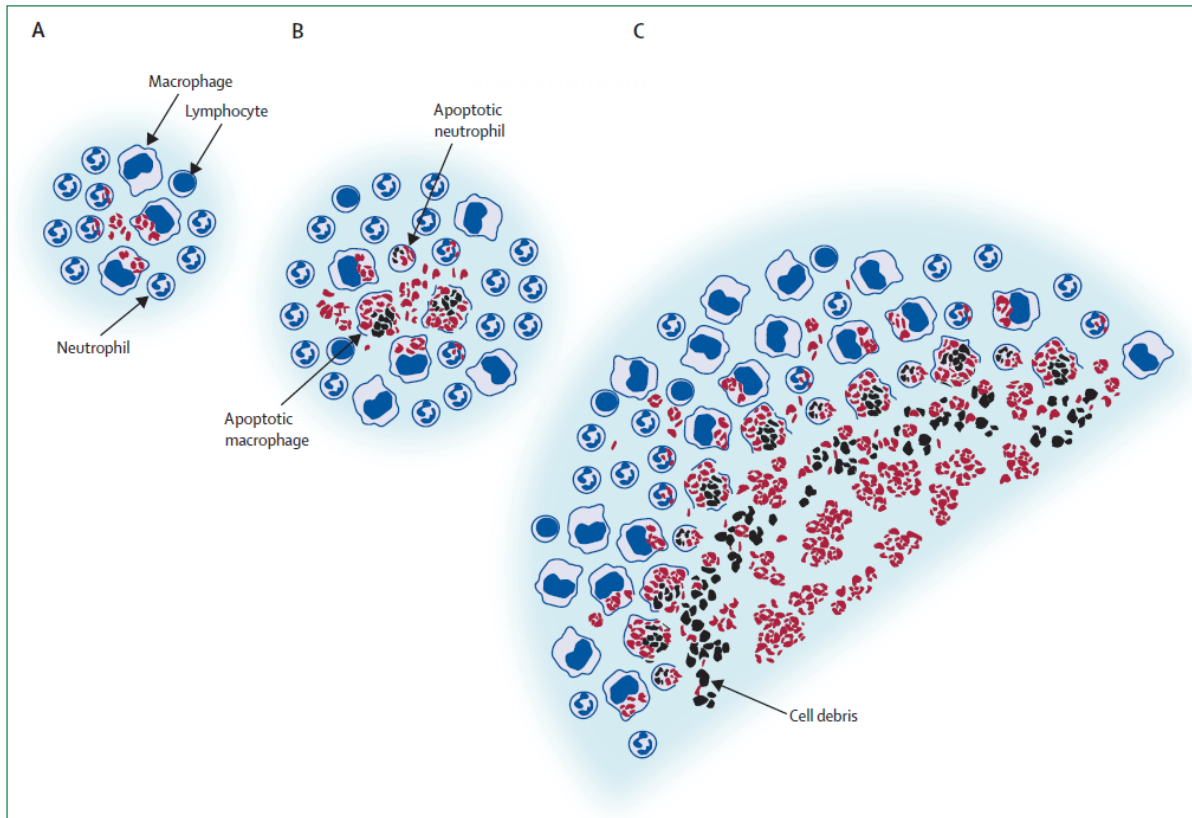
Les mycobactéries sont généralement considérées comme un prototype des pathogènes intracellulaires facultatifs (146). La prédominance extracellulaire de *M. ulcerans* sur les coupes histologiques a posé des questions fondamentales sur la biologie du bacille, le type de réponse immunitaire qu'il induit, et donc le modèle de vaccination le plus probablement efficace à lui opposer. Il semble néanmoins clair que la bactérie possède un stade de développement intracellulaire, comme le suggère des observations histologiques, immunologiques et cliniques.

Différentes observations attestent en effet d'une invasion intracellulaire par la bactérie chez la souris et l'homme, tout au long de l'infection (109). *In vitro*, la mycobactérie se multiplie dans un modèle de macrophages murins, et *in vivo* on observe des macrophages au cytoplasme saturé de mycobactéries, suggérant également une multiplication intracellulaire (Figure 9) (147). Sur le plan histologique chez l'homme, les lésions en voie de guérison révèlent un infiltrat inflammatoire important, la formation de granulomes géantocellulaires sans nécrose caséuse et la présence intracellulaire de la bactérie (148). La réponse immunitaire adaptative contre *M. ulcerans* fait intervenir essentiellement l'immunité cellulaire et présente un profil TH<sub>1</sub> qui est plus généralement associée aux germes intracellulaires (147). Il a été montré *via* un modèle de souris invalidées pour le gène de l'interféron-gamma que cette cytokine permet le contrôle de l'infection par une souche de *M. ulcerans* à condition que celle-ci ne produise pas de mycolactone. La sécrétion de mycolactone

inhiberait l'acidification des phagolysosomes et pourrait représenter un nouveau mécanisme d'évasion immunitaire par un pathogène intracellulaire (149). Enfin, sur le plan clinique, une réaction d'hypersensibilité retardée, marqueur d'une immunité adaptative cellulaire ciblant un germe intracellulaire, ici provoquée par l'injection intradermique d'un sonicat de *M. ulcerans*, la buruline, est observée selon les séries chez 70 à 100% des patients (23,150,151).

L'identification de bacilles intracellulaires et d'une réponse immunitaire de type cellulaire aboutissant à la formation de granulomes et à une hypersensibilité retardée sont autant d'arguments pour considérer *M. ulcerans* comme un pathogène intracellulaire facultatif. Sa physiologie est néanmoins unique suite à l'acquisition du plasmide pMUM et à la production de mycolactone, molécule aux multiples propriétés fascinantes (cytotoxicité, immunomodulation, analgésie, peut-être antiseptie). Ces propriétés apparaissent parfois contradictoires pour une compréhension de la biologie de *M. ulcerans* (par exemple, inhibition de la phagocytose et cytotoxicité compatibles avec un mode de vie extracellulaire vs. inhibition de l'acidification du phagolysosome compatible avec un mode de vie intracellulaire). Il faut garder en mémoire que la concentration et la cinétique de production de mycolactone dans les lésions est inconnue *in natura*, ce qui rend périlleux une interprétation physiopathologique fine du rôle de la mycolactone.





**Figure 12. Modèle physiopathologique de l'infection à *M. ulcerans***

- (A) Stade précoce de l'infection : phagocytose de *M. ulcerans* par des macrophages et des neutrophiles
- (B) Stade intermédiaire : poursuite de la phagocytose; apoptose des cellules phagocytaires précoces; production de mycolactone par les bacilles extracellulaires
- (C) Stade avancé, typique des lésions anatomopathologiques diagnostiques : au centre de la lésion nécrotique, grand nombre de bacilles extracellulaires; en marge de la nécrose, infiltrat inflammatoire comprenant des bacilles intracellulaires

*Adapté de (109)*

En conclusion, le modèle physiopathologique actuellement privilégié de l'infection à *M. ulcerans* est celui d'une multiplication intramacrophagique de la bactérie, suivie par une lyse cellulaire et la production de mycolactone cytotoxique et immunosuppressive provoquant une nécrose sous-cutanée étendue et limitant la réponse immunitaire, puis d'une invasion secondaire de nouveaux phagocytes (Figure 12). Ce modèle permet d'expliquer la présence de larges zones nécrotiques avec une abondance de bactéries extracellulaires sans infiltrat inflammatoire coexistant avec en périphérie la multiplication intracellulaire de la bactérie dans un faible nombre de phagocytes (109). Le rôle de la réponse humorale n'est pas encore définitivement établi. Plusieurs études ont observé une production d'anticorps contre *M. ulcerans*. De manière intéressante, ces anticorps discriminent qualitativement les individus vivant en et hors zone d'endémie, mais quantitativement en zone d'endémie il est difficile de distinguer les atteints et les non-atteints. On peut interpréter ces observations comme un signe d'infection latente ou comme un signe d'immunité protectrice chez certains individus exposés non-atteints (17,151–154).

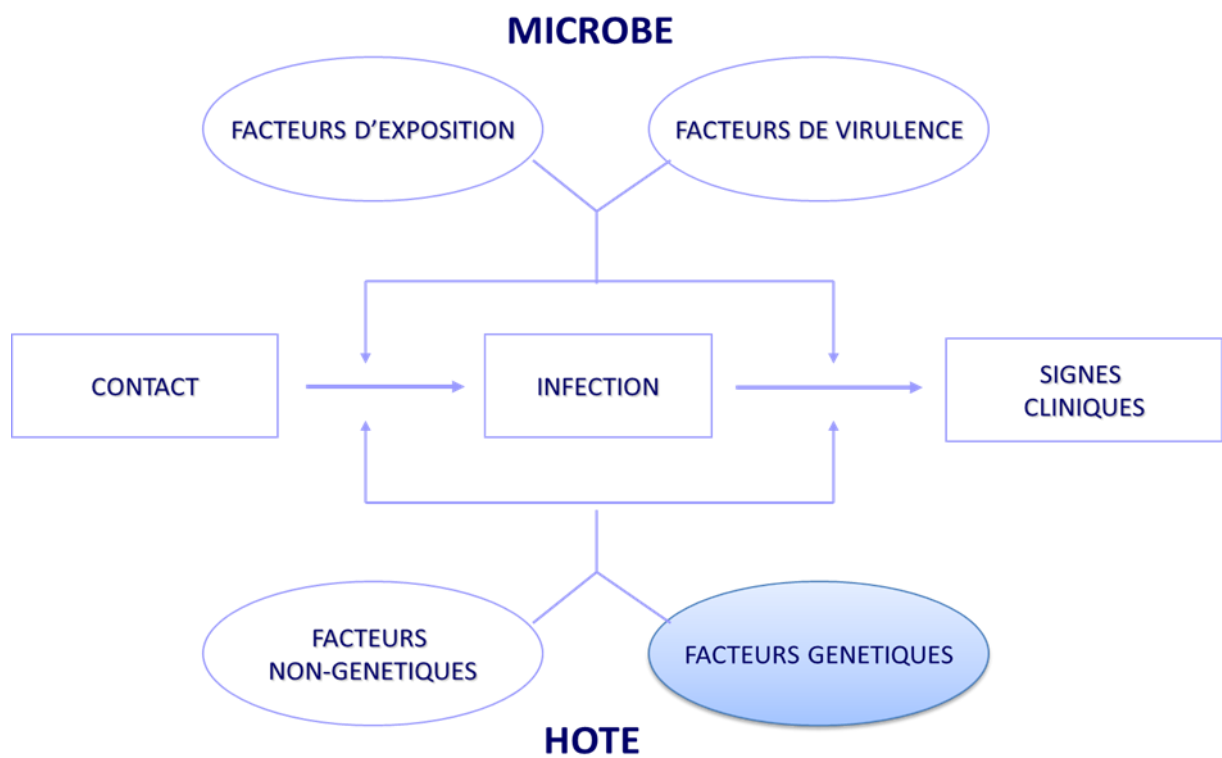


Figure 13. Du contact avec *M. ulcerans* au développement de signes cliniques :

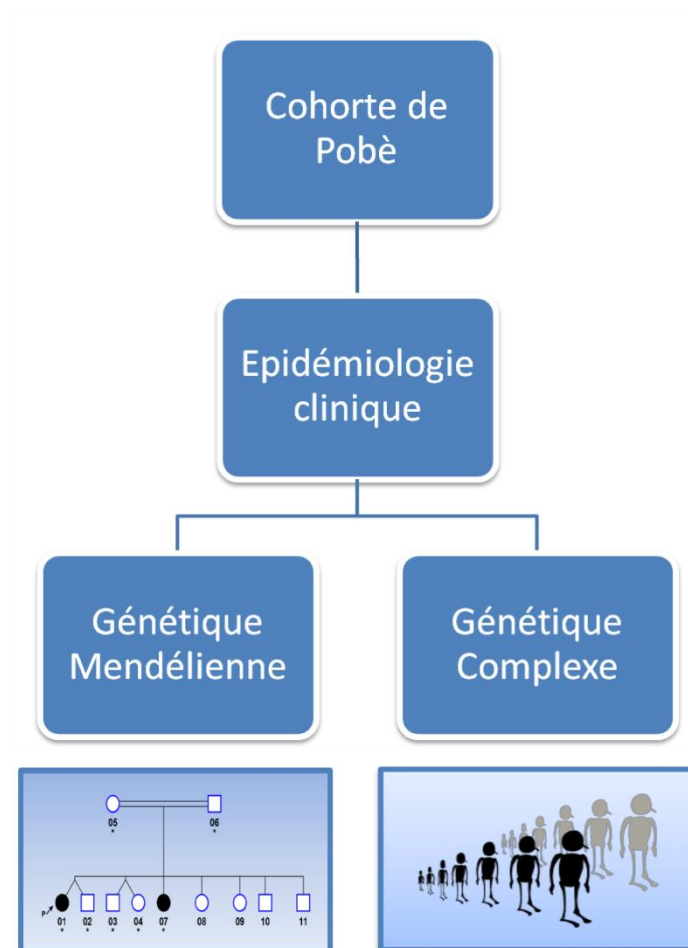
facteurs de variabilité

*adapté de (155)*

## Exploration de l'épidémiologie et de la génétique humaine de l'ulcère de Buruli : motivations et objectifs

On constate une variabilité majeure dans le devenir des individus exposés à *M. ulcerans* tant sur le plan clinique que cellulaire : (i) dans les zones d'endémie où l'exposition à la mycobactérie est chronique et intense, la majorité des individus ne développe pas de lésion ; (ii) parmi les individus développant une lésion, on observe un large spectre de manifestations cliniques depuis la lésion cutanée ulcéreuse unique jusqu'à l'ostéomyélite sévère disséminée ; (iii) les exposés non-atteints et les atteints ne développent pas tous une réaction d'hypersensibilité retardée à l'injection intradermique de buruline, avec respectivement des taux de positivité de l'ordre de 25 et 90% (23); (iv) l'histopathologie de la maladie au diagnostic montre la formation d'un granulome chez seulement 50% des patients au stade d'ulcère, et après traitement l'évolution histologique vers un infiltrat inflammatoire majeur avec formations de granulomes n'est pas systématique (148,156); (v) le profil de réponse cytokinique à l'infection n'est pas homogène (98); (vi) la production d'anticorps anti-*M. ulcerans* est également hautement variable tant chez les exposés que les atteints (17,152).

Cette variabilité interindividuelle peut être expliquée par des facteurs liés à la bactérie et d'autres liés à l'hôte (Figure 13, adaptée de (155)). Le premier facteur est le facteur d'exposition bactérienne, tant qualitatif (être exposé ou pas; mode d'exposition : abrasion superficielle ou inoculation sous-cutanée par piqûre) que quantitatif (charge bactérienne initiale). Un second facteur est la virulence bactérienne (type de mycolactone en particulier). Ceci étant dit, à l'échelle d'un pays et *a fortiori* à l'échelle d'un foyer hyperendémique, la souche de *M. ulcerans* semble extrêmement stable (157). Au Ghana et au Bénin, une étude comparative de 13 souches de *M. ulcerans* n'a identifié que 400 positions variables sur l'ensemble du génome de la bactérie (125). Sur le plan de l'hôte, des facteurs non spécifiquement génétiques peuvent modifier l'efficacité de la réponse immunitaire, comme l'âge, l'état nutritionnel, ou une immunodéficience acquise. Nous faisons l'hypothèse que cette variabilité interindividuelle est également au moins partiellement sous contrôle génétique de l'hôte.



**Figure 14. Schéma de l'étude**

La constitution d'une base de données des cas d'UB soignés au Centre de Dépistage et de Traitement de l'UB de Pobè au Bénin a fait l'objet d'une étude épidémiologique descriptive et analytique. Une famille présentant une forme particulièrement sévère de d'UB a été exploré par une approche de génétique mendélienne. Plus généralement, la susceptibilité génétique complexe à l'UB a été étudiée par étude d'association pangénomique de plan cas-témoin.

En faveur de cette hypothèse, une étude rapporte qu'un antécédent d'UB dans la famille multiplie par 5.5 (IC 95% [3.0-10.0]) le risque de développer un UB (37). Il n'y a pas d'études comparant les taux de concordance entre jumeaux mono- et dizygotes dans l'UB ni d'étude évaluant l'héritabilité de l'UB. Une étude d'association préliminaire sur gène-candidat suggère un rôle de certains polymorphismes du gène *SLC11A1* (précédemment nommé *NRAMP1*) dans le contrôle de la maladie dans une population ghanéenne, sans réplique (158). Les données animales montrent que différentes lignées de souris ne présentent pas la même susceptibilité à *M. ulcerans* dans des conditions expérimentales. L'exemple le plus caricatural est certainement la souris invalidée pour le gène produisant le TNF $\alpha$ , chez laquelle *M. ulcerans* se multiplie significativement plus rapidement que chez une souris "sauvage" (159). Enfin, un argument indirect mais probablement plus convaincant encore est le rôle majeur joué par les facteurs génétiques de l'hôte dans les mycobactérioses communes, la tuberculose et la lèpre, et rares, comme les infections disséminées à BCG (160–165).

Cette étude a pour objectif d'identifier les bases épidémiologiques et génétiques de l'infection à *M. ulcerans* chez l'homme (Figure 14). Tout le travail de cette thèse est fondé sur la collaboration entre l'équipe d'épidémiologie génétique du laboratoire de Génétique Humaine des Maladies Infectieuses dirigée par Laurent Abel à l'Hôpital Necker-Enfants Malades à Paris, l'équipe Biologie et Ecologie de *M. ulcerans* dirigée par Laurent Marsollier au Centre Hospitalo-Universitaire d'Angers, et l'équipe soignante dirigée par Annick Chauty au CDTUB de Pobè au Bénin. Elle a bénéficié du soutien et de financements de l'INSERM, de l'Agence Nationale de la Recherche, de la Fondation pour la Recherche Médicale, de l'association Raoul Follereau France, et de l'Institut Imagine.



**Figure 15. Le lieu de l'étude : Pobè, région de Ouémé-Plateau, Bénin**

(A) Pobè est situé en zone rurale au sud du Bénin

(B) Bâtiments d'hospitalisation du CDTUB

(C) A. Chauty, A. Alcaïs, Q. Vincent au CDTUB (juin 2010)

(D-F) Environs de Pobè : piste, maisons en torchis, activités domestiques et baignade en bord de fleuve

Le CDTUB de Pobè, financé par l'association Raoul Follereau France, a été fondé en 2003 et emploie environ cinquante personnes, dont trois médecins et un chirurgien. Il est situé en zone rurale à l'est du Bénin, près de la frontière nigériane (Figure 15). La capacité d'hospitalisation est de 58 lits. Le bloc opératoire fonctionne deux fois par semaine. Le CDTUB est doté d'un laboratoire d'analyses permettant la réalisation des principaux examens biologiques. Les radiologies sont effectuées à proximité à l'hôpital de zone de Pobè. Le CDTUB coordonne les soins de l'UB en hospitalisation et en ambulatoire à Pobè et dans les 15 postes avancés de soins au cœur des villages endémiques dans un rayon d'environ 30 kilomètres. L'organisation des soins prévoit trois tournées médicales hebdomadaires dans les villages endémiques. Le nombre annuel de patients traités est de l'ordre de 200. Tous les patients sont revus à distance du traitement pour la détection des rechutes et l'évaluation des séquelles fonctionnelles. L'équipe dirigée par Laurent Marsollier au CHU d'Angers, spécialiste de la biologie et de l'écologie de *M. ulcerans*, collabore depuis de nombreuses années avec le CDTUB de Pobè. Le diagnostic PCR de la maladie est fait chaque semaine à Angers sur des prélèvements en provenance du Bénin. Depuis environ deux ans, l'équipe de Laurent Marsollier organise le transfert local de cette technologie nécessitant du matériel et des compétences de pointe en biologie moléculaire. Le diagnostic PCR est actuellement réalisé en doublon à Angers et à Pobè.

C'est dans ce contexte que j'ai entrepris d'explorer l'épidémiologie et la génétique humaine de l'ulcère de Buruli. Cette thèse s'articule autour d'un premier axe épidémiologique descriptif et analytique de la plus grande cohorte de cas confirmés d'UB au monde, d'un second axe de génétique mendélienne qui explore une famille consanguine ayant présenté une forme exceptionnellement sévère de la maladie et d'un troisième axe de génétique complexe ayant pour objectif de mettre en évidence les facteurs de risque génétiques de l'UB dans un contexte cas-témoin.





# Epidémiologie descriptive et analytique de l'ulcère de Buruli

---

**Les publications issues de cette partie sont reproduites en Annexe :**

**Vincent QB**, Ardant M-F, Adeye A, Goundote A, Saint-André J-P, Cottin J, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health*. jul 2014;2(7):e422 - e430.

**Vincent QB**, Ardant M-F, Marsollier L, Chauty A, Alcaïs A. HIV infection and Buruli ulcer in Africa. *Lancet Infect Dis*. September 2014;14(9):796–7.

Pommelet V, **Vincent QB**, Ardant M-F, Adeye A, Tanase A, Tondeur L, et al. Findings in Patients From Benin With Osteomyelitis and Polymerase Chain Reaction-Confirmed *Mycobacterium ulcerans* Infection. *Clin Infect Dis*. July 21, 2014; epub ahead of print.

**Vincent QB**, Ardant M-F, Marsollier L, Chauty A, Alcaïs A. Defining and targeting high-risk populations in Buruli ulcer—Authors' reply. *Lancet Glob Health*. 2014 Nov;2(11):e630.

**Un extrait des questions des relecteurs scientifiques mandatés par le Lancet Global Health pour l'évaluation de notre manuscrit est également reproduit en Annexe, avec nos réponses.**



## Epidémiologie de l'UB : contexte

Plusieurs études séminales ont construit les bases actuelles de l'épidémiologie de l'UB. La première étude d'envergure date de 1971 avec la description de l'épidémie d'UB du camp de réfugiés de Kinyara en Ouganda (27). Le déplacement d'une population de 2500 réfugiés initialement indemnes en zone d'endémie a abouti au développement d'un UB chez 10% d'entre eux. Cette étude prospective, au plan quasi-expérimental, a fourni les premières estimations de la période d'incubation de l'UB, de son incidence en fonction de l'âge et du sexe (incidence maximale chez les moins de 15 ans), et de l'influence de la proximité entre l'habitation et le fleuve sur l'incidence de la maladie. Au Ghana, en 1999, une étude transversale nationale a été conduite pour dénombrer le nombre de cas actifs et de cas guéris d'UB dans l'ensemble du pays, et a identifié plus de 5000 patients (15). On y trouve une estimation de la prévalence de la maladie, la distribution des patients selon leur âge et leur sexe et une description de la localisation des lésions. L'étude de Zagnanado au Bénin a rassemblé entre 1630 et 2399 cas d'UB entre 1997 et 2003 (18,30,41). Elle montre en particulier que l'incidence de l'UB dépasse celle de la lèpre et de la tuberculose dans cette région, et s'intéresse également à l'incidence de l'UB par âge et par sexe, ainsi qu'à la localisation des lésions. Elle identifie une courbe d'incidence bimodale de l'UB, avec une plus grande incidence chez les moins de 15 ans et les 70-74 ans. Une correspondance rapporte brièvement la distribution de l'âge, du sexe, du type et de la localisation des lésions chez 2598 patients issus de quatre centres au Bénin entre 2003 et 2005 (29). Enfin, une étude de 750 cas au Ghana rapporte spécifiquement la localisation des lésions (31). D'autres études ont rapporté un nombre plus modeste de cas (100 à 300 patients), souvent dans le cadre d'études cas-témoins visant à identifier les facteurs de risque d'UB (proximité à l'eau stagnante, soin des plaies, port de vêtements couvrants, usage de moustiquaire) (16,19,23,33,34,38,39).

Chacune de ces études a contribué à établir le tableau épidémiologique de l'UB. Néanmoins, ces études souffrent de limitations significatives, malgré des tailles d'échantillon apparemment grandes : certaines sont purement descriptives ; d'autres reposent sur une grande part de diagnostics rétrospectifs faits d'après l'observation d'une cicatrice évocatrice d'UB ; d'autres reposent sur des diagnostics prospectifs cliniques mais non confirmés par un examen de laboratoire ; et certaines ont un effectif réel largement amoindri par de fortes proportions de données manquantes pour des analyses clés. L'OMS insiste sur la confirmation diagnostique des cas suspects d'UB par un ou plusieurs examens de laboratoire afin d'éliminer des diagnostics différentiels comme les ulcères phagédéniques, la tuberculose cutanée et certaines mycoses profondes (82,83). Le test le plus sensible et le plus spécifique est la détection de l'ADN de *M. ulcerans* par PCR dans les lésions prélevées par aspiration à l'aiguille fine, écouvillonnage ou biopsie (57,82,83). La coloration de Ziehl-Neelsen, la culture bactérienne ou l'analyse anatomopathologique peuvent également être utilisées. Dans les plus grandes séries sur l'UB (>750 cas), les diagnostics étaient confirmés par un examen de laboratoire au mieux dans environ 50% des cas (15,27,29–31).

Dans ce travail, nous rapportons l'étude et le suivi d'une large cohorte de cas d'UB traités au Bénin, confirmés par un examen de laboratoire, et décrivons la diversité des présentations cliniques de l'UB, leur variation en fonction de l'âge et du sexe, et leur impact sur la survenue de séquelles fonctionnelles. Dans un deuxième temps, nous explorons le lien potentiel entre infection par le VIH et présentation clinique de l'UB. Ces travaux ont fait l'objet de plusieurs publications reproduites en Annexe (166,167). Notre étude a suscité un commentaire stimulant auquel nous avons pu répondre (Annexe, publication imminente). Certaines questions particulièrement éclairantes des relecteurs scientifiques (reviewers) ont également été reproduites en Annexe. Par ailleurs, un aspect particulier de notre étude sur les ostéomyélites à *M. ulcerans* a généré une hypothèse de travail explorée de manière approfondie par Virginie Pommelet dans l'équipe d'épidémiologie des maladies émergentes dirigée par Arnaud Fontanet à l'Institut Pasteur (168).

## Epidémiologie de l'UB : population et méthodes

Les données cliniques et les résultats d'examens de laboratoire ont été collectés chez tous les cas consécutifs suspects d'UB entre 2005 et 2011 au CDTUB de Pobè, Bénin.

Nous avons recueilli prospectivement l'âge au diagnostic, le sexe, l'origine géographique et le statut VIH du patient ; la forme clinique (nodule, plaque, œdème, ulcère, ostéomyélite), la catégorie OMS de taille (diamètre maximal <5cm, 5-15 cm, ≥15 cm) et le site de la lésion (membre supérieur/inférieur droit/gauche, thorax, abdomen, tête, périnée) ; la ou les méthodes de confirmation diagnostique (culture, coloration de Ziehl-Neelsen, PCR *IS2404*) ; le caractère médical et/ou chirurgical de la prise en charge et le délai de guérison. Il est important de garder à l'esprit qu'un patient d'UB peut souffrir de lésions à un ou plusieurs sites (ce qui définit la multifocalité, par exemple bras droit et gauche) et peut présenter plusieurs types de lésions à un seul site (par exemple, un ulcère sur plaque). Les patients de notre cohorte ont également été suivis pour déterminer la survenue de séquelles fonctionnelles, depuis la perte de 10° ou plus d'amplitude articulaire jusqu'à l'amputation d'un membre. La partie supérieure du corps était définie comme la tête, le thorax, l'abdomen et les membres supérieurs ; et la partie inférieure du corps comme les membres inférieurs et le périnée. L'accès aux dossiers et aux données informatisées a été approuvé par le comité éthique institutionnel du CDTUB et par les autorités du Plan National de Lutte contre l'UB.

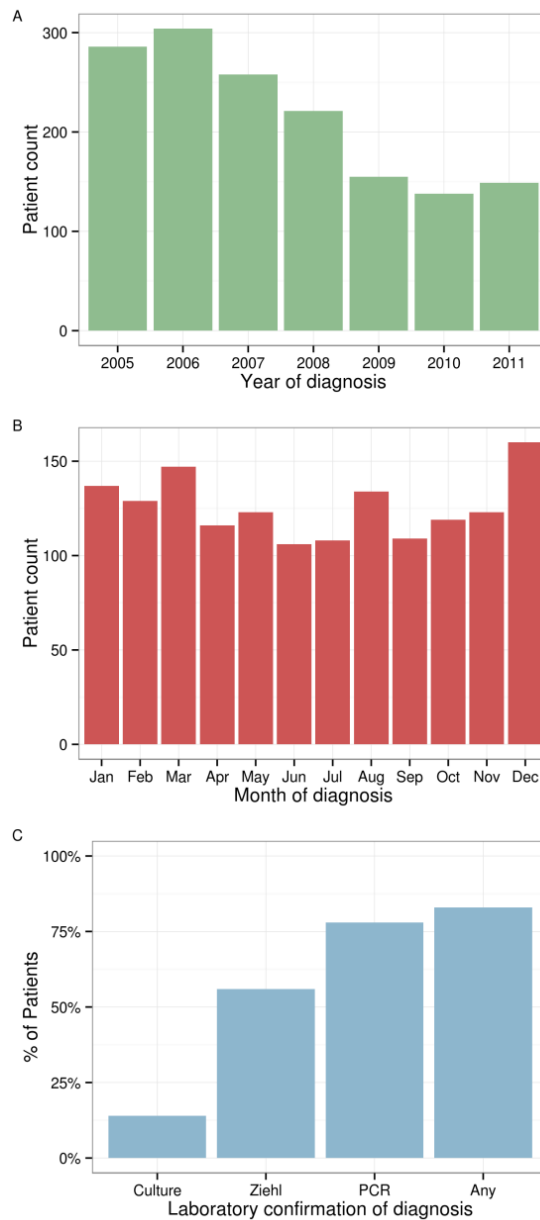
En accord avec les recommandations OMS concernant l'importance d'une confirmation diagnostique de l'UB par au moins un examen complémentaire (82), nous avons réalisé nos analyses sur les cas confirmés uniquement. Nous avons décrit la présentation clinique de l'UB et analysé l'impact de l'âge au diagnostic et du sexe sur cette présentation. Dix caractéristiques de la présentation et de l'évolution clinique ont été évaluées : être un homme, présenter une lésion sur la partie supérieure du corps, présenter un nodule, une plaque, un ulcère, une forme oedémateuse, une ostéomyélite, présenter une lésion de plus de 15 centimètre de diamètre maximum, présenter plusieurs lésions (multifocalité), et présenter des séquelles fonctionnelles fixées. Nous avons appliqué systématiquement trois modèles de régression logistique, considérant respectivement l'âge, le sexe, et l'âge et le sexe comme variables explicatives de la présence de chaque élément mentionné ci-dessus.

Nous avons ensuite testé l'effet potentiel de chaque forme clinique d'UB sur la survenue de séquelles fonctionnelles dans des modèles univariés et multivariés. Une présentation clinique sévère d'UB a été définie comme une présentation clinique comportant au moins un élément clinique à risque de séquelles fonctionnelles. Nous avons validé cette définition en mesurant l'impact de la sévérité sur le délai de guérison (test de Wilcoxon).

L'âge a été analysé comme une variable continue par la méthode des polynômes fractionnels (p-values correspondantes notées « MFP-p ») (169). Pour faciliter l'interprétation des résultats, un odds-ratio (OR) pour l'âge dichotomisé avec un seuil de 15 ans est également rapporté (OR correspondant noté  $OR_{15}$ ). L'estimation des paramètres du modèle et les tests d'hypothèse ont été réalisés dans le cadre de la théorie statistique du maximum de vraisemblance. Les analyses ont été réalisées avec le logiciel statistique R (fonctions *glm*), et les paquets additionnels *mfp* et *ggplot2* (170–172).

Dans un deuxième temps, nous avons exploré l'impact de l'infection par le VIH sur l'incidence et la présentation clinique de l'UB. Les chiffres de prévalence nationale du VIH ont été obtenus sur le site du programme UNAIDS des Nations-Unies (173). Le statut VIH des patients était déterminé par deux tests immunologiques. Etant donné le faible nombre total de coinfections, nous avons d'abord réalisé l'analyse sur l'ensemble des cas d'UB traités par le CDTUB, quel que soit leur statut de confirmation diagnostique par un examen de laboratoire, puis sur le groupe des cas d'UB confirmés par PCR.





**Figure 16. Distribution des cas d'UB en fonction de l'année du diagnostic (A), du mois du diagnostic (B) et du statut de confirmation par examen complémentaire (C)**

**Pobè, Bénin (ensemble des cas déclarés à l'OMS, 2005-2011, n=1511)**

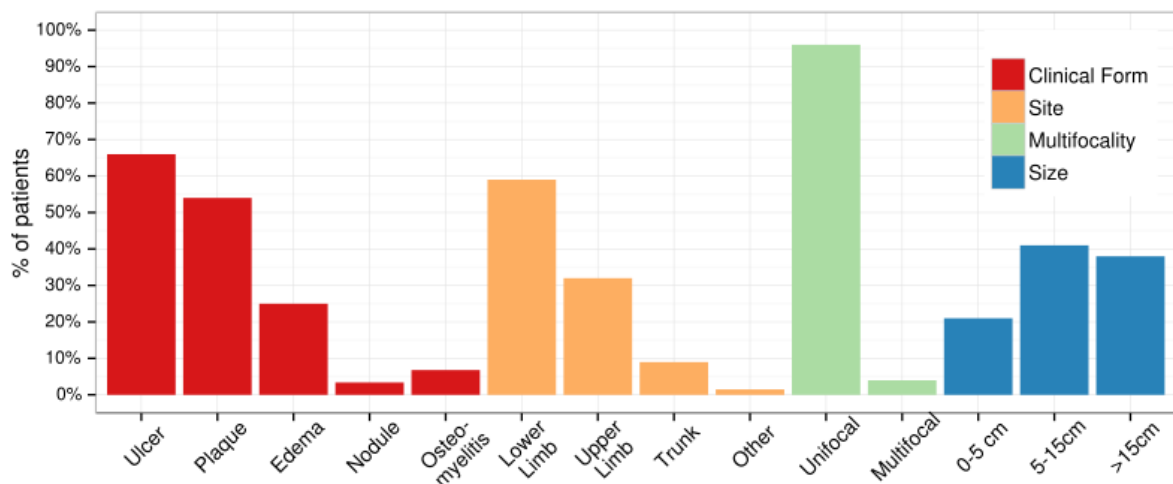
## Epidémiologie de l'UB : résultats

### Diagnostic et prise en charge de l'UB au CDTUB de Pobè

Un total de 1511 cas a été traité pour un UB entre 2005 et 2011 au CDTUB de Pobè, Bénin. Le nombre annuel de cas était maximum en 2006 (n=304), avec une baisse graduelle à 150 cas par an environ de 2009 à 2011 (Figure 16A). Le nombre annuel de cas en provenance du Nigéria dont la frontière est proche du CDTUB a augmenté entre 2005 (n=5) et 2011 (n=19). La distribution des cas n'était globalement pas régulière tout au long de l'année, avec plus de cas diagnostiqués pendant la saison sèche (Novembre-Mars) que pendant la saison des pluies (Avril-Octobre) ( $p=0.014$ ) (Figure 16B).

Le diagnostic était confirmé par au moins un examen complémentaire dans 83% des cas (n=1251) et par deux examens complémentaires dans 54% des cas (n=812). Les cas n'ayant pas bénéficié de la réalisation d'un examen complémentaire représentaient 3.3% de la cohorte (n=50). La PCR était positive dans 78% des cas (n=1177), la coloration de Ziehl-Neelsen dans 56% des cas (n=842) et la culture dans 14% des cas (n=214) (Figure 16C). En ce qui concerne les ostéomyélites, 50% des cas ont été confirmés sur la lésion osseuse, directement sur un échantillon osseux chirurgical (43%) ou sur un liquide de fistulisation (7%). Les autres cas correspondaient soit à des lésions fermées contigües à une lésion osseuse, confirmés par aspiration à l'aiguille fine (11%), ou des ulcères contigus à une lésion osseuse, confirmés par écouvillonnage (24%) ou biopsie cutanée (15%).

96% des patients (n=1452) ont reçu une antibiothérapie (streptomycine et rifampicine dans 91% des cas, n=1377) et 56% des patients (n=839) ont eu recours à la chirurgie (du débridement simple à l'excision-greffe). Soixante-neuf patients ont été traités avec succès par un traitement antibiotique oral exclusif par clarithromycine et rifampicine (2007-2011). Le délai médian de guérison était de 88 jours. Douze cas ont souffert de nouvelles lésions d'UB plus d'un an (entre 1.5 et 5 ans) après guérison de l'épisode initial.



**Figure 17. Présentation clinique de l'UB :  
forme clinique, site atteint, nombre et taille des lésions**

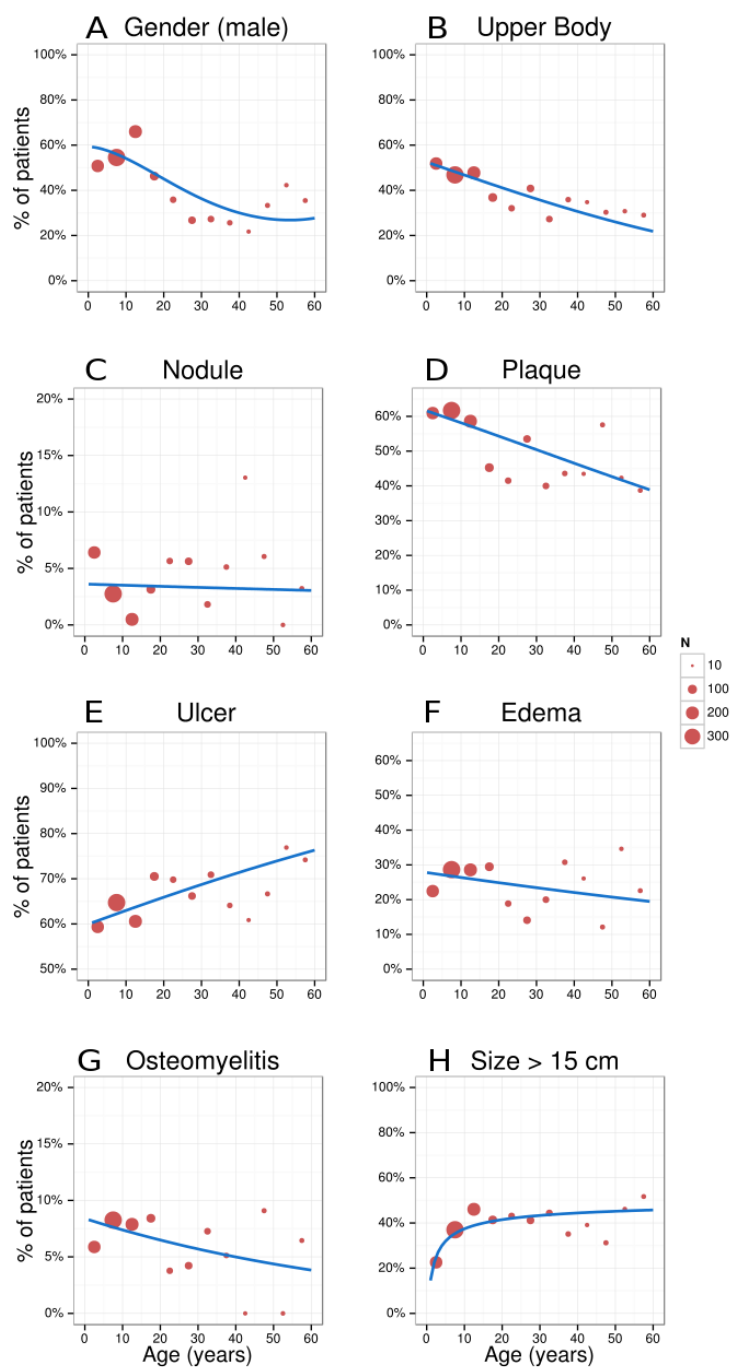
**Pobè, Bénin (cas confirmés par laboratoire, 2005-2011)**

A noter qu'un patient d'UB peut souffrir de lésions à un ou plusieurs sites (ce qui définit la multifocalité, par exemple bras droit et gauche) et peut présenter plusieurs types de lésions à un seul site (par exemple, un ulcère sur plaque).

## **La présentation clinique la plus classique de l'UB est une ulcération large d'un membre inférieur**

Pour décrire de façon précise la présentation clinique de l'UB, la variation de cette présentation en fonction de l'âge et du sexe, et l'impact de cette présentation sur la fixation de séquelles fonctionnelles, nous avons restreint les analyses aux cas d'UB confirmés par laboratoire sans coinfection VIH (n=1227).

Il ressort de ces analyses qu'un patient d'UB typique est un enfant présentant un large ulcère unique sur un membre inférieur (Figure 17). En effet, l'âge médian au diagnostic était de 12 ans (étendue interquartile 7-28 ans, moyenne 19.3 ans). Une comparaison avec l'âge médian de la population béninoise (18 ans en 2010, (174)) suggère que l'UB est surreprésenté chez les enfants. Les lésions étaient larges, puisque environ 40% des patients (n=444) présentaient une lésion de plus de 15 centimètres de diamètre. Plus de 95% des patients (n=1172) présentaient une lésion unique. La lésion siège sur les membres inférieurs dans 60% des cas (n=733). Les ulcères étaient la forme clinique la plus fréquente (66% des patients, n=805), suivis par les plaques (54%, n=668), les formes oedémateuses (25%, n=307) et les nodules (3%, n=42). Une ostéomyélite était observée dans 7% (n=82) des cas, dont au minimum 17% (n=14) ne présentait pas de lésion cutanée d'UB active ou résolue.



**Figure 18. Présentation clinique de l'UB :  
impact de l'âge au diagnostic**

**Pobè, Bénin (cas confirmés par laboratoire, 2005-2011)**

La proportion de patients présentant chaque trait exploré est présentée en fonction de l'âge. En rouge, les cercles indiquent la proportion observée par groupes d'âge de 5 ans ([0-5[... [55-60[). La surface de chaque cercle est proportionnelle à l'effectif du groupe d'âge (voir échelle « N »). A titre d'exemple, il y avait ~200 patients dans le groupe des 10-15 ans, dont 47% présentaient des lésions sur la partie supérieure du corps (Panneau B). En bleu, est représentée la proportion de patients prédite par régression logistique en modélisant l'âge comme une variable continue par la méthode des polynômes fractionnels. Les formules exactes des modèles sont en Annexe. Le nombre d'individu total est de 1163 pour l'étendue des lésions et de 1224 pour tous les autres éléments considérés.

## **Le sexe masculin est surreprésenté parmi les cas pédiatriques d'UB tandis que le sexe féminin est surreprésenté parmi les cas adultes d'UB**

En considérant l'ensemble des patients, le sex-ratio était équilibré (sexe féminin 52%, n=640 ; sexe masculin 48%, n=587, p=0.13). Une distorsion majeure du sex-ratio était néanmoins observée en fonction de l'âge, avec une prédominance masculine chez les jeunes patients et une prédominance féminine chez les patients plus âgés. Les patients masculins représentaient en effet 57% (n=427) des patients âgés de moins de 15 ans, mais seulement 33% (n=158) des patients âgés de plus de 15 ans (Tableau 3,  $OR_{15}=2.59$ , IC95% [2.04-3.30]; MFP-p<0.0001). Cet effet est le plus fort de ceux recensés dans le Tableau 3 et n'est pas attribuable à la démographie de la population béninoise dont le sex-ratio est équilibré aussi bien chez les moins de 15 ans que chez les plus de 15 ans. Autrement dit, l'âge au diagnostic était très différent chez les patients masculins et féminins, avec un âge médian (moyen) au diagnostic de 10 (16) ans chez les hommes et de 15 (22.5) ans chez les femmes (p<0.0001).

**Tableau 3 Impact de l'âge et du sexe sur la présentation clinique de l'UB**

Caractéristique clinique	% (n) avec la caractéristique clinique			Analyse univariée <sup>a</sup>			Analyse bivariée <sup>b</sup>		
	Variable explicative			OR	95% CI	p-value <sup>c</sup>	OR	95% CI	p-value
Sexe masculin	Age	continue	-	-	-	<10 <sup>-4</sup>	-	-	-
		> 15	33.5% (n=158)	1	-	-	-	-	-
		≤ 15	56.7% (n=427)	2.59	[2.04-3.30]	<10 <sup>-4</sup>	-	-	-
Localisation sur la partie supérieure du corps	Age	continue	-	-	-	<10 <sup>-4</sup>	-	-	<10 <sup>-4</sup>
		> 15	31.8% (n=150)	1	-	-	1	-	-
		≤ 15	48.3% (n=364)	2.00	[1.58-2.55]	<10 <sup>-4</sup>	2.13	[1.67-2.74]	<10 <sup>-4</sup>
	Sexe	Féminin	43.0% (n=275)	1	-	-	1	-	-
		Masculin	40.7% (n=239)	0.91	[0.73-1.14]	0.4241	0.78	[0.61-0.98]	0.0366
Nodule	Age	continue	-	-	-	0.7496	-	-	0.4886
		> 15	4.0% (n=19)	1	-	-	1	-	-
		≤ 15	3.1% (n=23)	0.75	[0.40-1.41]	0.3639	0.85	[0.45-1.62]	0.6247
	Sexe	Féminin	4.4% (n=28)	1	-	-	1	-	-
		Masculin	2.4% (n=14)	0.53	[0.27-1.01]	0.0528	0.51	[0.26-0.99]	0.0427
Plaque	Age	continue	-	-	-	<10 <sup>-4</sup>	-	-	<10 <sup>-4</sup>
		> 15	44.8% (n=211)	1	-	-	1	-	-
		≤ 15	60.7% (n=457)	1.90	[1.51-2.40]	<10 <sup>-4</sup>	2.06	[1.62-2.63]	<10 <sup>-4</sup>
	Sexe	Féminin	56.4% (n=361)	1	-	-	1	-	-
		Masculin	52.3% (n=307)	0.85	[0.68-1.06]	0.1491	0.75	[0.60-0.95]	0.0167
Ulcère	Age	continue	-	-	-	3.7 x 10 <sup>-4</sup>	-	-	<10 <sup>-4</sup>
		> 15	70.7% (n=333)	1	-	-	1	-	-
		≤ 15	62.3% (n=469)	0.68	[0.53-0.88]	0.0024	0.64	[0.49-0.82]	4.2 x 10 <sup>-4</sup>
	Sexe	Féminin	63.3% (n=405)	1	-	-	1	-	-
		Masculin	68.1% (n=400)	1.24	[0.98-1.57]	0.0731	1.36	[1.07-1.74]	0.0120
Oedème	Age	continue	-	-	-	0.0445	-	-	0.0413
		> 15	21.9% (n=103)	1	-	-	1	-	-
		≤ 15	27.1% (n=204)	1.33	[1.01-1.75]	0.0390	1.35	[1.02-1.78]	0.0349
	Sexe	Féminin	25.0% (n=160)	1	-	-	1	-	-
		Masculin	25.0% (n=147)	1.00	[0.77-1.30]	0.9863	0.95	[0.73-1.24]	0.7214
Ostéomyélite	Age	continue	-	-	-	0.0560	-	-	0.1993
		> 15	5.3% (n=25)	1	-	-	1	-	-
		≤ 15	7.6% (n=57)	1.46	[0.91-2.41]	0.1181	1.23	[0.76-2.05]	0.4106
	Sexe	Féminin	4.4% (n=28)	1	-	-	1	-	-
		Masculin	9.2% (n=54)	2.21	[1.39-3.59]	6.8 x 10 <sup>-4</sup>	2.09	[1.30-3.38]	0.0019
Etendue ≥ 15cm	Age	continue	-	-	-	<10 <sup>-4</sup>	-	-	<10 <sup>-4</sup>
		> 15	41.8% (n=189)	1	-	-	1	-	-
		≤ 15	35.9% (n=255)	0.78	[0.61-0.99]	0.0422	0.74	[0.57-0.94]	0.0157
	Sexe	Féminin	36.3% (n=226)	1	-	-	1	-	-
		Masculin	40.1% (n=218)	1.18	[0.93-1.49]	0.1746	1.28	[1.00-1.63]	0.0466
Atteinte multifocale	Age	continue	-	-	-	0.1292	-	-	0.1308
		> 15	3.8% (n=18)	1	-	-	1	-	-
		≤ 15	4.7% (n=35)	1.23	[0.69-2.24]	0.4878	1.22	[0.68-2.26]	0.5020
	Sexe	Féminin	4.2% (n=27)	1	-	-	1	-	-
		Masculin	4.4% (n=26)	1.05	[0.60-1.82]	0.8654	0.97	[0.55-1.70]	0.9114
Forme Sévère	Age	continue	-	-	-	2.7 x 10 <sup>-4</sup>	-	-	1.6 x 10 <sup>-4</sup>
		> 15	52.1% (n=245)	1	-	-	1	-	-
		≤ 15	49.4% (n=371)	0.90	[0.71-1.13]	0.3537	0.85	[0.67-1.08]	0.1736
	Sexe	Féminin	48.0% (n=306)	1	-	-	1	-	-
		Masculin	52.9% (n=310)	1.22	[0.97-1.53]	0.0842	1.27	[1.01-1.59]	0.0416
Survenue de séquelles fonctionnelles	Age	continue	-	-	-	0.1098	-	-	0.2006
		> 15	20.1% (n=79)	1	-	-	1	-	-
		≤ 15	23.1% (n=150)	1.20	[0.88-1.63]	0.2479	1.13	[0.83-1.55]	0.4425
	Sexe	Féminin	19.8% (n=109)	1	-	-	1	-	-
		Masculin	24.3% (n=120)	1.30	[0.97-1.75]	0.0783	1.26	[0.93-1.70]	0.1289

a. Analyse univariée de chaque caractéristique clinique en fonction des variables explicatives âge ou sexe. Les odds-ratios bruts (OR), leur intervalle de confiance à 95% (CI) et la p-value du test de rapport de vraisemblance sont rapportés.

b. Analyse bivariée de chaque caractéristique clinique en fonction des variables explicatives âge et sexe. Les odds-ratios ajustés (OR), leur intervalle de confiance à 95% (CI) et la p-value du test de rapport de vraisemblance sont rapportés.

c. L'analyse de l'effet de l'âge a été conduite selon deux approches: en considérant l'âge comme une variable continue modélisée par polynômes fractionnels (MFP) et considérant l'âge comme une variable catégorielles avec un seuil à l'âge de 15 ans (la Figure 18 donne une visualisation des MFP ; les détails des MFP sont en Annexe).

## La présentation clinique de l'UB varie en fonction de l'âge et du sexe

L'ensemble des résultats concernant la variation de la présentation clinique en fonction de l'âge et du sexe est donné en Tableau 3 et Figure 18. Les résultats principaux sont résumés ici.

Neuf pour cent (n=54) des patients masculins ont développé une ostéomyélite contre 4% (n=28) des patients féminins (OR=2.21, 95%CI [1.39-3.59],  $p=0.0007$ ). Cet effet du sexe sur la présentation clinique est l'effet le plus marqué du Tableau 3 après la distorsion du sex-ratio avec l'âge.

L'impact de l'âge sur la présentation clinique était très significatif en ce qui concerne la localisation des lésions, le développement d'une plaque et le développement d'un ulcère. En effet, les patients les plus jeunes étaient particulièrement susceptibles de développer des lésions sur la partie supérieure du corps (OR<sub>15</sub>=2.00, 95%CI [1.58-2.55]; MFP- $p<0.0001$ , Figure 18B) et de présenter une plaque (OR<sub>15</sub>=1.90, 95%CI [1.51-2.40]; MFP- $p<0.0001$ , Figure 18D). A l'inverse, les patients les plus âgés étaient particulièrement susceptibles de présenter une ulcération (OR<sub>15</sub>=0.68, 95%CI [0.53-0.88]; MFP- $p=0.0004$ , Figure 18E).

Plusieurs associations s'approchaient de la significativité statistique. Les nodules étaient plus fréquents chez les patients féminins (OR=0.53, 95%CI [0.27-1.01],  $p=0.05$ ). Les ostéomyélites étaient plus fréquentes chez les jeunes patients (OR<sub>15</sub>=1.46, 95%CI [0.91-2.41]; MFP- $p=0.06$ , Figure 18G).



**Tableau 4. Impact de la présentation clinique de l'UB sur la survenue de séquelles à long-terme**

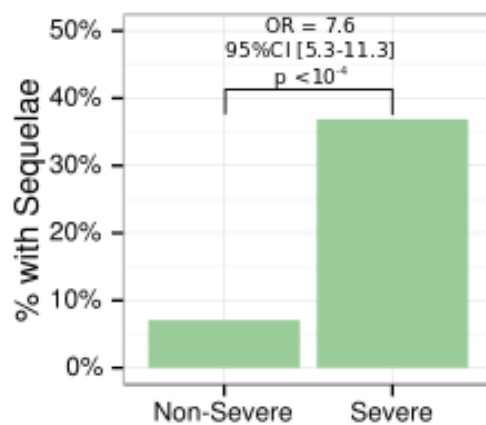
**Pobè, Bénin (cas confirmés par laboratoire, 2005-2011)**

			Analyse univariée <sup>a</sup>			Analyse multivariée <sup>b</sup>		
Caractéristique clinique		% (n) avec des séquelles fonctionnelles	OR	95% CI	p-value	OR	95% CI	p-value
Localisation sur la partie supérieure du corps	Non	20.5% (n=124)	1	-	-	1	-	-
	Oui	23.9% (n=105)	1.22	[0.91-1.63]	0.1930	1.59	[1.10-2.30]	0.0128
Nodule	Absence	22.6% (n=227)	1	-	-	1	-	-
	Présence	5.4% (n=2)	0.20	[0.03-0.65]	0.0045	0.47	[0.07-1.73]	0.2829
Plaquette	Absence	30.6% (n=143)	1	-	-	1	-	-
	Présence	15.0% (n=86)	0.40	[0.29-0.54]	<10 <sup>-4</sup>	0.64	[0.45-0.92]	0.0156
Ulcère	Absence	19.5% (n=69)	1	-	-	1	-	-
	Présence	23.2% (n=160)	1.24	[0.91-1.71]	0.1758	2.07	[1.33-3.28]	0.0012
Oedème	Absence	18.9% (n=146)	1	-	-	1	-	-
	Présence	30.6% (n=83)	1.89	[1.38-2.59]	<10 <sup>-4</sup>	1.39	[0.90-2.15]	0.1347
Ostéomyélite	Absence	19.0% (n=185)	1	-	-	1	-	-
	Présence	65.7% (n=44)	8.18	[4.87-14.1]	<10 <sup>-4</sup>	6.48	[2.27-20.9]	0.0004
Etendue	< 15 cm	7.6% (n=47)	1	-	-	1	-	-
	≥ 15 cm	40.6% (n=153)	8.30	[5.82-12.0]	<10 <sup>-4</sup>	6.95	[4.70-10.4]	<10 <sup>-4</sup>
Atteinte multifocale	Non	20.9% (n=210)	1	-	-	1	-	-
	Oui	48.7% (n=19)	3.59	[1.87-6.87]	0.0002	0.91	[0.37-2.18]	0.8328
Forme sévère	Non	7.1% (n=37)	1	-	-			
	Oui	36.9% (n=192)	7.64	[5.29-11.3]	<10 <sup>-4</sup>			

- a. Analyse univariée de la survenue de séquelles fonctionnelles à long-terme en fonction de la présentation clinique. Les odds-ratios bruts (OR), leur intervalle de confiance à 95% (CI) et la p-value du test de rapport de vraisemblance sont rapportés.
- b. Analyse multivariée de la survenue de séquelles fonctionnelles à long-terme en fonction de l'ensemble des items de la présentation clinique (hors sévérité). Les odds-ratios ajustés (OR), leur intervalle de confiance à 95% (CI) et la p-value du test de rapport de vraisemblance sont rapportés.

## **Certaines présentations cliniques concentrent le risque de séquelles fonctionnelles et définissent la sévérité de l'UB**

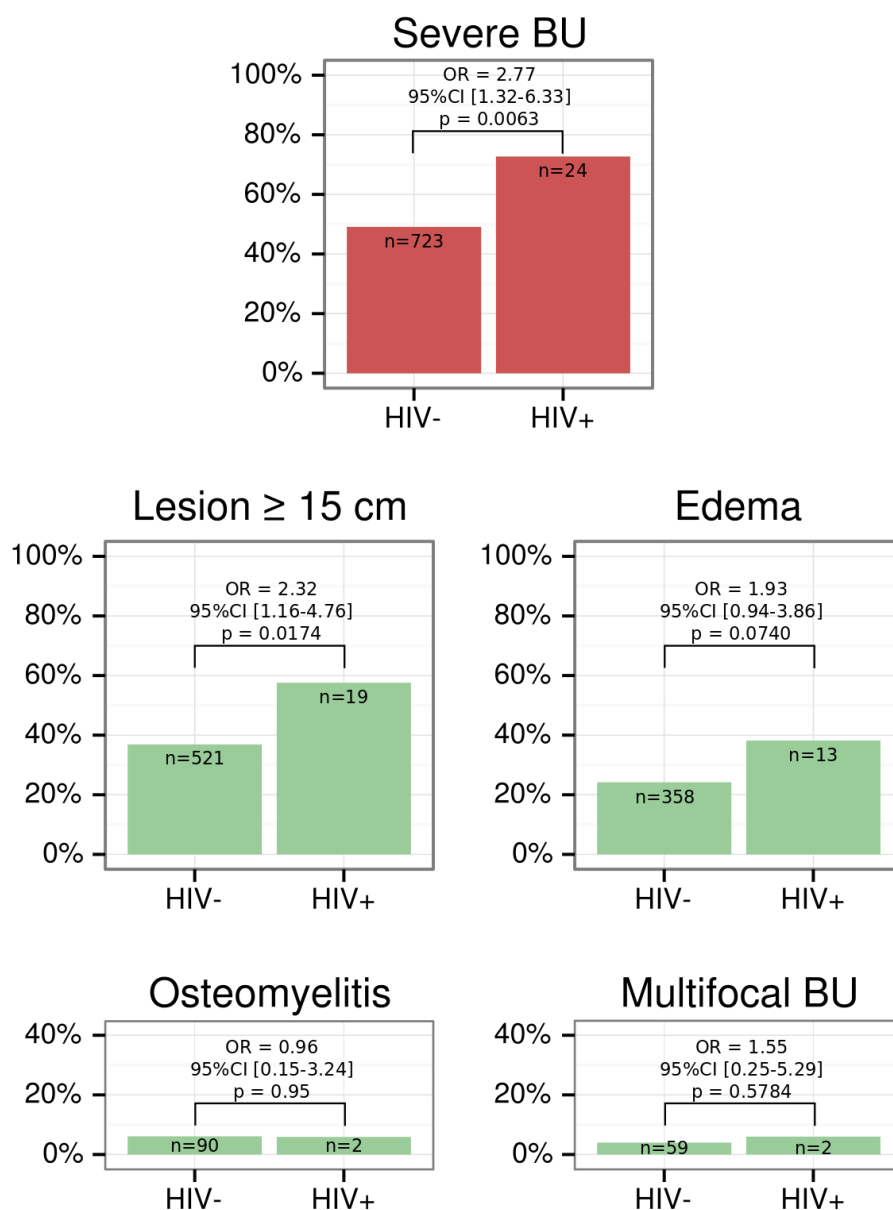
Un suivi à long-terme (médiane de 359 jours) était disponible chez 85% des patients (n=1043). Des séquelles fonctionnelles fixées étaient observées chez 22% (n=229) d'entre eux. 1% (n=10) des patients a subi une amputation, emportant une phalange jusqu'à un membre entier. L'incidence de séquelles fonctionnelles ne variait pas significativement en fonction de l'âge ou du sexe des patients (Tableau 3). En revanche, les patients présentant une forme oedémateuse, une forme osseuse, une lésion étendue (plus de 15 centimètres de diamètre) ou plusieurs lésions (forme multifocale) étaient à risque significativement plus élevé de séquelles (Tableau 4). Cette constatation nous a conduits à proposer une définition opérationnelle de la sévérité d'un cas d'UB comme une présentation clinique comportant l'une de ces quatre caractéristiques cliniques. Les cas sévères d'UB représentaient 50% (n=616) des patients. Globalement, 37% (n=192) des patients présentant un UB sévère ont développé des séquelles fonctionnelles contre 7% (n=37) des patients présentant un UB non-sévère (Figure 19, OR=7.64, 95%CI [5.29-11.31],  $p<0.0001$ ). L'analyse multivariée a révélé des corrélations entre les quatre composantes cliniques de la sévérité, puisque seules les présentations d'une forme osseuse ou d'une lésion étendue conservaient des effets significatifs indépendants (OR=6.48, 95%CI [2.27-20.9],  $p=0.0004$  and OR=6.95, 95%CI [4.70-10.4],  $p<0.0001$ , respectivement). Le délai médian de guérison était de 81 jours dans les formes non-sévères et de 107 jours dans les formes sévères ( $p<0.0001$ ).



**Figure 19. Survenue de séquelles à long-terme chez les patients en fonction de la sévérité de la présentation clinique de l'UB**

**Pobè, Bénin (cas confirmés par laboratoire, 2005-2011)**

Un cas sévère d'UB était défini comme une présentation clinique comportant une forme oedémateuse, une forme osseuse, plusieurs lésions (multifocalité) ou une lésion étendue (plus de 15 cm).



**Figure 20. Sévérité de la présentation clinique de l'UB en fonction du statut VIH**

**Pobè, Bénin (ensemble des cas déclarés à l'OMS, 2005-2011)**

Un cas sévère d'UB était défini comme une présentation clinique comportant une forme oedémateuse, une forme osseuse, plusieurs lésions (multifocalité) ou une lésion étendue (plus de 15cm).

## **L'infection par le VIH est associée à la sévérité de la présentation clinique de l'UB**

Sur l'ensemble des 1511 cas au diagnostic clinique d'UB, une séropositivité VIH était identifiée chez 2.3% (n=34) des patients et 5% (n=25) des patients âgés de 15 à 49 ans (n=500). Cette proportion est significativement plus élevée que la prévalence nationale de la séropositivité chez les 15-49 ans, estimée à 1.1% en 2012 au Bénin ( $p < 0.0001$ ) (173). Plus de 70% (n=24) des patients séropositifs ont développé une forme sévère d'UB, contre 50% (n=723) des patients séronégatifs (OR=2.77, 95%CI [1.32-6.33],  $p=0.006$  ; Figure 20). L'analyse du sous-groupe des cas d'UB confirmés par PCR confirme cette observation (OR=2.59, 95% CI [1.06-7.27],  $p=0.037$ ). L'impact de la séropositivité sur la sévérité de la présentation clinique de l'UB était essentiellement expliqué par une plus grande fréquence des lésions étendues et/ou oedémateuses chez les patients coinfectés (Figure 20). L'ajustement sur l'âge et le sexe ne modifiait pas les résultats.

## Epidémiologie de l'UB : discussion

Malgré sa progression rapide depuis les années 1980, l'UB demeure une maladie infectieuse négligée. Le CDTUB de Pobè est actuellement à la pointe de la recherche clinique et fondamentale sur cette maladie (29,75,96,100,117,152,175). Notre étude épidémiologique bénéficie de plusieurs caractéristiques fondamentales dont l'association est sans précédent dans le contexte de l'UB : (i) il s'agit d'une étude représentative de l'Afrique de l'Ouest où l'immense majorité des cas d'UB sont recensés, (ii) le nombre de cas inclus est élevé, (iii) tous les cas ont été diagnostiqués prospectivement et confirmés par au moins un examen complémentaire (une PCR dans 95% des cas), (iv) la fixation de séquelles fonctionnelles a été évaluée à long-terme chez plus de 85% des patients, (v) la collecte prospective des données a été minutieuse aboutissant à une faible proportion de données manquantes. Nous avons décrit la présentation clinique de l'UB, sa variation en fonction de l'âge et du sexe des patients et son impact sur la survenue de séquelles fonctionnelles. Nous avons également évalué l'impact du VIH sur l'infection par *M. ulcerans*.

Plusieurs caractéristiques épidémiologiques de l'UB rapportées régulièrement par les études antérieures ont été confirmées, ce qui est en faveur de la généralisabilité de nos résultats : l'UB est une maladie essentiellement pédiatrique (âge médian au diagnostic de 12 ans) ; les lésions sont très majoritairement unifocales (>90%) (19,27,29–31) ; elles prédominent sur les membres inférieurs (~60%) (15,23,29) ; l'ulcère est la forme clinique la plus courante (~70%) (29,31) ; une proportion significative de patients présente une ostéomyélite (>5%) (29,30) ; les séquelles fonctionnelles sont fréquentes (>20%) (23,34). Notre étude rapporte de nombreux autres résultats, dont six paraissent particulièrement importants et seront discutés ici : la variation du sex-ratio avec l'âge ; la surreprésentation des garçons parmi les cas d'ostéomyélites ; l'existence d'ostéomyélites « exclusives » chez des patients sans lésion cutanée d'UB active ou passée ; la forte corrélation entre la présentation clinique et la fixation de séquelles fonctionnelles ; la sous-estimation de l'incidence de l'UB en Afrique ; et l'impact de la coinfection par le VIH sur la maladie.

Le sex-ratio global est équilibré dans l'UB, mais varie significativement en fonction de l'âge au diagnostic. Les patients masculins représentent près de 60% des patients âgés de moins de quinze ans, mais un tiers seulement des patients de plus de quinze ans. La variation du sex-ratio avec l'âge au diagnostic d'UB a été rapportée précédemment (15,27,33–35), mais de manière surprenante, cette information a toujours été omise ou même niée dans les revues consacrées à l'UB (57,58,83,86,109,176). Un dimorphisme sexuel est fréquemment observé dans les maladies infectieuses (pour une revue détaillée, voir (177)). Dans le contexte des mycobactérioses humaines, les infections rares par *M. marinum* ou *M. avium intracellulare* et les infections fréquentes par *M. leprae* et *M. tuberculosis* sont plus fréquentes chez les hommes adultes que les femmes adultes (126,178–181). Les hypothèses explicatives couvrent un large spectre de possibilités, depuis une exposition différentielle au pathogène en fonction du sexe jusqu'à une réponse immunitaire spécifique à chaque sexe. Des études expérimentales d'inoculation de *M. marinum* chez la souris ont comparé l'évolution clinique et microbiologique chez la souris mâle, la souris femelle et la souris mâle castrée pour démontrer que des niveaux élevés de testostérone diminuaient la résistance à cette bactérie (182). Dans notre étude observationnelle, *in natura*, l'infection par *M. ulcerans* suivait une tendance opposée chez les adultes. Ceci illustre peut-être la physiopathologie unique de *M. ulcerans* conférée par la sécrétion d'une toxine lipidique, la mycolactone. Des études spécifiques sont nécessaires pour comprendre les mécanismes sous-jacents à la dynamique du sex-ratio en fonction de l'âge dans l'UB.

Suite à notre publication, la prédominance de l'UB dans la population pédiatrique et la variation du sex-ratio avec l'âge ont à nouveau été confirmées sur des données camerounaises analysées avec une méthodologie différente de la nôtre (publication imminente). Les auteurs de ce travail confirmaient nos résultats mais argumentaient que nos conclusions étaient peut-être excessives dans la mesure où nous n'avions pas normalisé la distribution de l'âge au diagnostic par les données démographiques béninoises pour obtenir des incidences standardisées. Nous avons pu commenter la différence entre nos approches en réponse à cette correspondance (Annexe,

publication imminente). De manière intéressante, un des relecteurs (reviewers) de notre manuscrit avait déjà discuté de la normalisation des données par les effectifs de la population générale. Si cette normalisation est nécessaire pour savoir si l'UB est surreprésenté chez les enfants, il nous semble qu'elle est ne l'est pas pour savoir si le sex-ratio est équilibré dans l'UB, dans la mesure où le sex-ratio au Bénin est de 1 pour toutes les tranches d'âge (174). Par ailleurs, nous pouvons noter que cette correction n'est pas pertinente pour la majorité des questions abordées dans notre étude, telle que la surreprésentation d'une classe d'âge ou de sexe pour certaines formes cliniques d'UB au sein des patients UB. Nous fournissons en annexe les réponses à cette question et à plusieurs autres questions éclairantes des relecteurs scientifiques.

Dans notre étude, le risque de développer une ostéomyélite à *M. ulcerans* était significativement plus élevé chez les patients masculins (OR=2.21, 95%CI [1.39-3.59]) et tendait à être plus élevé les jeunes patients. De manière intéressante, l'effet de l'âge sur le risque d'ostéomyélite était significatif si l'on augmentait la taille de l'échantillon en considérant l'ensemble des patients, quel que soit leur confirmation diagnostique par un examen de laboratoire. Une précédente étude n'avait pas détecté d'effet de l'âge et du sexe sur le développement d'ostéomyélites (123). A la lumière de nos résultats, ceci pourrait refléter un manque de puissance de cette étude, puisque la puissance de détection d'un OR de 2.2 dans un échantillon de 186 patients masculins et 187 patients féminins, pour un niveau d'erreur de type I de 5%, était inférieure à 40%. Etant donné ses implications cliniques et physiopathologiques, notre découverte mérite d'être confirmée dans une étude indépendante de forte puissance. L'atteinte préférentielle des garçons dans l'ostéomyélite est néanmoins rapportée dans le contexte général des ostéomyélites hors UB, ce qui conforte notre résultat (183). Les garçons atteints d'UB doivent d'ores et déjà être surveillés avec une attention particulière pour détecter précocement la survenue d'une ostéomyélite. Comprendre le mécanisme par lequel les garçons sont plus à risque de développer une ostéomyélite est d'un intérêt scientifique majeur et devrait faire l'objet d'études dédiées dans le futur.



Nous rapportons pour la première fois des cas d'ostéomyélite exclusive chez des patients sans lésion cutanée d'UB active ou passée, qui représentent 17% des ostéomyélites à *M. ulcerans*. Nous avons appliqué des critères stringents pour définir une ostéomyélite exclusive, ce qui implique que cette estimation représente plutôt une borne inférieure. C'est d'ailleurs la conclusion de l'étude spécifique réalisée par Virginie Pommelet qui évalue que 25% des ostéomyélites sont exclusives (168). Ces patients sont venus consulter avec une ostéomyélite clinique d'aspect non-spécifique et l'infection par *M. ulcerans* n'a été identifiée qu'à travers les examens complémentaires. Un biais de mémoire pourrait expliquer que certains patients ne se soient pas rappelés d'une ancienne lésion d'UB de taille limitée, qui n'aurait pas laissé de cicatrice décelable par l'équipe soignante, mais tous les patients ne sont probablement pas dans ce cas. Combien même ce serait le cas, notre observation resterait d'une grande pertinence clinique, épidémiologique et physiopathologique.

La pertinence clinique est évidente : *M. ulcerans* est un candidat étiologique devant une ostéomyélite dans une zone endémique d'UB, ce qui implique de faire des prélèvements à la recherche de la bactérie et une adaptation thérapeutique éventuelle. En ce qui concerne la physiopathologie, ces cas exclusifs sont la preuve que l'ostéomyélite dans l'UB ne résulte pas nécessairement de la multiplication incontrôlée de la bactérie dans son tissu d'élection, la peau. Certains individus sont susceptibles à une invasion osseuse par une quantité modérée de bactéries sans destruction tissulaire contigüe. Il est important de noter que nous ne remettons pas en cause le mode de transmission de *M. ulcerans* : nous n'affirmons pas que les patients n'ont pas été piqués ou mordus par un insecte vecteur, ou qu'une abrasion cutanée n'a pas été infectée par contact avec de l'eau contaminée. Nous rapportons que certains patients développent des lésions osseuses d'UB en l'absence d'un réservoir cutané massif pour la mycobactérie. Etant donné la clonalité de la bactérie à l'échelle locale (125,157), cette observation pose des questions sur la source de la variabilité interindividuelle de la réponse immunitaire à ce pathogène. Il s'agit d'un point important qui pourrait être exploré par différentes études, en particulier génétiques. Enfin, cette observation est en faveur d'un mécanisme de dissémination hématogène de la bactérie depuis le site d'inoculation.

Nous avons évalué de manière systématique l'impact de la présentation clinique sur la fixation de séquelles fonctionnelles. Quatre éléments de la présentation clinique de l'UB étaient associés à un délai de guérison significativement plus long et étaient à risque de séquelles fonctionnelles. Nous proposons donc une définition opérationnelle d'un cas sévère d'UB comme une évolution clinique comportant une forme oedémateuse, une forme osseuse, une lésion étendue ( $\geq 15$  cm) ou des lésions multiples (multifocalité). L'intensité de l'association entre la sévérité de la présentation clinique de l'UB telle que définie ici et le développement de séquelles fonctionnelles (OR=7.6) est d'une significativité clinique incontestable. Les patients présentant l'un de ces quatre types de lésion doivent bénéficier de soins cliniques spécifiques, avec un suivi rapproché, une rééducation fonctionnelle intense et une prise en charge de reconstruction chirurgicale optimale. Notre observation soulève également la question de la raison pour laquelle certains patients développent une forme sévère d'UB. Le délai au diagnostic étant difficile à estimer, nous pouvons conjecturer qu'un délai diagnostique plus long expliquerait la sévérité de la présentation clinique chez certains patients. Certaines formes sévères, comme la forme oedémateuse, sont néanmoins d'évolution potentiellement rapide. L'identification de patients ayant connu une évolution rapide de l'UB vers une forme sévère serait intéressante pour explorer les variations interindividuelles de susceptibilité à l'UB.

Plusieurs éléments dans notre étude suggèrent une sous-estimation de l'incidence de l'UB en Afrique. Le CDTUB se trouve à quelques kilomètres de la frontière nigériane et accueille tous les ans des patients nigériens atteints d'UB. Alors que le nombre annuel de cas d'UB déclarés par le Nigéria à l'OMS est habituellement inférieur à dix (par exemple, 7 en 2010), une vingtaine de cas sont traités chaque année au CDTUB. La comparaison des superficies du Bénin et du Nigéria laisse à penser que la sous-estimation du nombre de cas d'UB dans ce pays est majeure. Par ailleurs, l'absence de diagnostic d'UB dans certaines présentations atypiques, comme les ostéomyélites exclusives, contribue également à la sous-estimation de l'incidence de l'UB.

L'interaction entre l'infection par le VIH et l'infection par *M. ulcerans* est un sujet controversé au centre des besoins de recherche actuels dans l'UB (53). On trouve dans la littérature des cas VIH+ présentant une forme sévère avec atteinte osseuse disséminée (43–45) mais également des formes d'évolution classique (46,47). Une étude cas-témoin au Bénin suggère une association VIH et UB, mais la proportion de VIH+ chez les témoins (0.3%, 2 individus sur 613) semble anormalement basse (48). Symétriquement, on trouve des cas VIH- d'évolution sévère (49–51) et une étude cas-témoin au Ghana ne retrouve pas d'association entre le VIH et l'UB (33). Notre étude montre que la coinfection VIH-UB est un événement clinique rare à Pobè, puisque 34 patients ont été identifiés ces 7 dernières années. Les patients coinfectés ont une présentation clinique d'UB plus fréquemment sévère que les patients VIH-. C'est également la conclusion d'une étude très récente au Cameroun où 83 patients coinfectés sont analysés (184). Nos données suggèrent par ailleurs une association entre séropositivité et risque d'UB, avec une séroprévalence VIH de plus de 5% chez les patients adultes contre 1.1% à l'échelle nationale. Cette observation, aussi faite au Cameroun, doit être interprétée prudemment, car notre étude et l'étude camerounaise ne fournissent pas le meilleur niveau de preuve en se basant sur des séroprévalences nationales et non un groupe de témoins locaux. L'étude camerounaise montre par ailleurs que la profondeur de l'immunosuppression est associée avec la sévérité de l'UB. Le décompte lymphocytaire des CD4 n'était malheureusement pas disponible dans notre étude.

Notre étude a également mis en lumière certains domaines où la collecte de données doit être plus précise pour différencier plusieurs situations, en particulier en ce qui concerne la guérison de l'UB. Il n'est pas évident de différencier une récurrence (une rechute infectieuse à court terme due à une population vivante de *M. ulcerans* non complètement éradiquée), une réaction paradoxale (une rechute non-infectieuse due à une réactivité immunologique excessive, peut-être liée à la levée de l'immunosuppression locale par la mycolactone), une réactivation (une rechute infectieuse à long-terme due à sortie de latence de *M. ulcerans*) et une réinfection (un épisode indépendant d'UB). Nous avons observé une dizaine de patients chez qui de nouvelles lésions d'UB sont apparues

plusieurs années après la guérison de l'épisode initial. Par analogie avec la tuberculose, il pourrait s'agir d'une réinfection ou d'une réactivation de bactéries en latence. Caractériser ces patients serait particulièrement important pour comprendre la réponse immunitaire à *M. ulcerans* dans l'optique du développement d'un vaccin (119).

La limite principale de notre étude est son caractère monocentrique. Celui-ci limite le nombre de patients pour certains sous-groupes (coinfection VIH-BU, ostéomyélites exclusives...). Il pose également la question de la généralisabilité de nos résultats. En revanche, il assure une homogénéité parfaite dans la collecte des données. Nous considérons que notre étude est représentative de l'UB dans les zones tropicales, où presque la totalité des cas d'UB survient. La publication quasiment simultanée de résultats concordants pour l'association UB-VIH au Bénin et au Cameroun nous conforte dans ce sens (167,184). De grandes séries de cas confirmés en provenance de zones non-tropicales, comme l'Australie et le Japon, seraient d'un intérêt particulier pour disséquer les contributions relatives de l'environnement, de l'exposition, du style de vie, des systèmes de santé, des souches microbiennes et de la génétique humaine dans le développement et la dynamique de l'UB chez l'homme.



# Genétique humaine de l'ulcère de Buruli : aspects mendéliens

---

**La publication issue de cette partie est reproduite en Annexe**

**Vincent QB** et al. L. Buruli ulcer of extreme severity maps to the beta-defensin locus on chromosome 8. To be submitted.



## Génétique mendélienne de l'UB : introduction

La constitution de la base de données concernant l'épidémiologie de l'UB au CDTUB, étudiée en première partie, a permis l'identification de la famille comportant le cas d'UB le plus sévère parmi plus de 1500 patients traités depuis 2003 dans cette zone hyperendémique (famille BU-A, Figure 21). Les parents se déclarent cousins issus de germains. Le cas index (07) a présenté à l'âge de 5 ans une forme œdémateuse d'UB, d'emblée multifocale, touchant les quatre membres, dont la dissémination ostéoarticulaire a progressé sous traitement, nécessitant une amputation. Sa sœur (01) a présenté à l'âge de 13 ans une forme œdémateuse intéressant l'ensemble du bras droit. Ces deux patientes étaient séronégatives pour l'infection par le VIH. La ségrégation de la maladie dans la famille, la consanguinité et la sévérité des cas sont compatibles avec la présence d'un défaut récessif dans un gène majeur de l'immunité contre *M. ulcerans*. Cette hypothèse est confortée par les découvertes de formes mendéliennes d'autres mycobactérioses communes et rares, comme la tuberculose et la bccgite disséminées (162,165,185).

Nous avons donc effectué trois types d'analyses génétiques complémentaires : une analyse de liaison génétique, une recherche de variation ponctuelle et une recherche de variation structurale, qui ont été rendues possibles par le génotypage de plus de 900 000 SNPs sur le génome des deux patientes, de trois frères et sœurs non-atteints, et des deux parents, ainsi que par le séquençage à haut-débit de l'ensemble des exons des gènes des deux patientes (exome).



## Génétique mendélienne de l'UB : sujets et méthodes

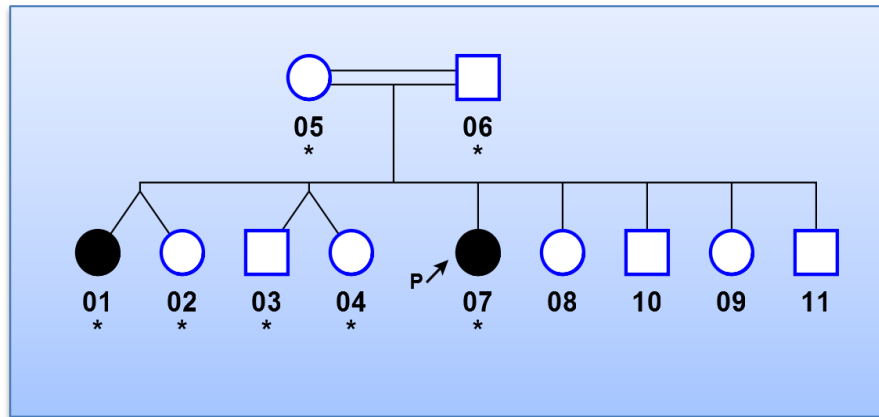
Le patient index a été identifié comme le cas d'UB le plus sévère parmi plus de 1500 patients traités au CDTUB de Pobè, Bénin, depuis sa fondation en 2003. Le diagnostic a été fait cliniquement et confirmé par coloration de Ziehl-Neelsen et amplification de la séquence spécifique IS2404 par PCR dans de multiples échantillons en provenance de multiples lésions, y compris osseuses. La sœur du patient index a également développé un UB sévère confirmé par PCR. Les parents et les 7 autres membres de la fratrie n'étaient pas atteints. Les parents étaient consanguins (cousins au second degré). La consanguinité dans cette famille suggère un mode de transmission récessif. Nous avons exploré cette famille par cartographie d'homozygotie (*homozygosity mapping*), une méthode d'analyse de liaison modèle-dépendante prenant en compte la consanguinité. La cartographie d'homozygotie identifie spécifiquement les régions du génome héritées de façon homozygote et identique par descendance par les individus atteints seulement. L'analyse de liaison a été réalisée par le logiciel MERLIN version 1.2, avec son option cluster permettant la prise en compte du déséquilibre de liaison (186,187). Un modèle récessif à pénétrance complète a été spécifié. La recherche sur la susceptibilité génétique de développer l'UB a été approuvée par le comité d'éthique du CDTUB (IRB00006860) et par les autorités du Plan National de Lutte contre l'UB. Un consentement éclairé a été collecté chez tous les membres de la famille participant à l'étude.

Une prise de sang a été réalisée chez les deux parents, non-atteints, chez trois enfants non-atteints et chez les deux enfants atteints (Figure 21A). L'ADN a été extrait du sang total par la méthode Nucleon BACC2 Genomic DNA Extraction (GE Healthcare), dosé par la méthode QuantIt Picogreen dsDNA (Life Technologies) et génotypé pour plus de 900,000 marqueurs génétiques, les Single Nucleotide Polymorphisms (SNPs) et pour plus de 900,000 marqueurs monomorphes (dédiés à l'identification de variants structuraux) sur la puce Affymetrix Genome-Wide 6.0. Des critères stringents de contrôle qualité ont été appliqués. Plus de détails sur le rationnel des contrôles qualité seront donnés dans la partie « Génétique complexe » de la thèse. Le contrôle qualité par individu consistait en un taux de génotypage supérieur à 95%, une compatibilité du sexe déclaré et du sexe

inféré génétiquement et une compatibilité des valeurs d'identité par état (IBS, *identity by state*) calculées par paires d'individus, avec le lien familial déclaré. Pour l'analyse de liaison, un contrôle qualité par SNP incluait un taux de génotypage intrafamilial de 100%, une fréquence intrafamiliale de l'allèle mineur non-nulle et l'absence d'erreur mendélienne. Plusieurs mesures de contrôle qualité classiques, comme la vérification respect de l'équilibre de Hardy-Weinberg, ne peuvent être appliquées à l'échelle d'une seule famille. Nous avons donc filtré les SNPs à une échelle populationnelle dans la population Hapmap Yoruba correspondante, et sélectionné les SNPs avec un taux de génotypage de 95% ou plus, une fréquence de l'allèle mineur non nulle, et une pvalue du test de l'équilibre de Hardy-Weinberg supérieure à 1% (188).

La cartographie d'homozygotie permet l'identification de régions contenant les variations causales, mais pas ces variations elles-mêmes. Nous avons donc cherché des variations structurales (larges insertions et délétions, plusieurs kilobases ou plus) par analyse du nombre de copies (Copy Number Variant, CNV) et des variations ponctuelles par séquençage de l'exome des deux patientes. L'identification de CNV repose sur l'analyse quantitative de l'intensité du signal d'hybridation des sondes nucléotidiques (le *logR ratio*, LRR) pour inférer un nombre normal, augmenté ou diminué de copies dans le génome du patient à la localisation de la sonde. Cette analyse CNV pangénomique a été réalisée avec le logiciel PennCNV version 2011Jun16 (189). Le séquençage de l'exome a été réalisé sur un séquenceur Illumina Genome Analyzer IIx (Illumina) après capture de l'exome par le kit SureSelect Human All Exon v2 kit (Agilent Technologies) couvrant 38 mégabases. Il s'agissait d'une technologie dite *single-end*. Les séquences obtenues étaient alignées avec le génome humain de référence (GRCh37) avec l'aligneur BWA (190). Les variants ont été déterminés à partir des séquences alignées par la suite Genome Analysis Toolkit (GATK) (191), par SAMtools (192), et Picard Tools (<http://picard.sourceforge.net>). Les options GATK UnifiedGenotyper et IndelGenotyperV2 ont respectivement permis de lire des variants de type substitution et insertion-délétion courte (quelques paires de base). Les variants avec une couverture de moins de 2x et une qualité Phred-Scaled de moins de 20 ont été filtrés.

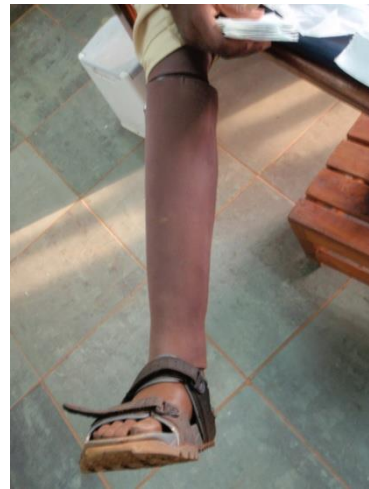
**A**



**B**



**C**



**D**



**E**



**Figure 21. Arbre généalogique et phénotype clinique de la famille BU-A**

(A) Généalogie de la famille, consanguine avec deux enfants atteints sur neuf. \*génomés. P: patient index

(B-C) Le cas index a subi une amputation et est maintenant appareillé

(D-E) Forme œdémateuse sévère chez la sœur et ses séquelles

YEAR	AGE	CLINICAL EVENTS
2000		<b>Birth</b>
2005	5 y.o	<b>Initial Diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: right thigh and leg; left foot</li> <li>- Septic Arthritis: right knee</li> <li>- Osteomyelitis : right fibula and tibia; left cuboid</li> <li>- Lab: Ziehl-Neelsen and PCR+ on both limbs</li> </ul> <b>Dissemination at 2 months post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: right arm and forearm</li> <li>- Septic Arthritis: right elbow</li> <li>- Osteomyelitis : right humerus and radius and ulna</li> <li>- Lab: Ziehl-Neelsen and PCR +</li> </ul> <div style="border: 1px solid red; padding: 5px; margin: 10px 0;"> <b>3 months post-diagnosis</b>            Amputation of the lower right limb         </div> <b>Dissemination at 6 months post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: left forearm</li> <li>- Osteomyelitis : left radius</li> <li>- Lab: Ziehl-Neelsen +; PCR not performed</li> </ul>
2006	6 y.o	<b>Dissemination at 1 year post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: left leg</li> <li>- Osteomyelitis : left fibula</li> <li>- Lab: Ziehl-Neelsen and PCR +</li> </ul>
2007	7 y.o	<b>Dissemination at 2 years post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: right arm</li> <li>- Osteomyelitis : right humerus</li> <li>- Lab: Ziehl-Neelsen +; PCR not performed</li> </ul>
2008	8 y.o	<b>Dissemination at 3 years post-diagnosis</b> <ul style="list-style-type: none"> <li>- Soft tissue collection: left calcaneal tendon</li> </ul>
2010	10 y.o	<b>Dissemination at 5 years post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: left leg</li> <li>- Osteomyelitis : left tibia</li> <li>- Lab: Ziehl-Neelsen and PCR +</li> </ul> <p><b>No further clinical event since 2010</b></p>

**Figure 22. Evolution clinique du patient index – 2005-2010**

Les quatre membres de la patiente ont progressivement été envahis par la mycobactérie.

## Génétique mendélienne de l'UB : résultats

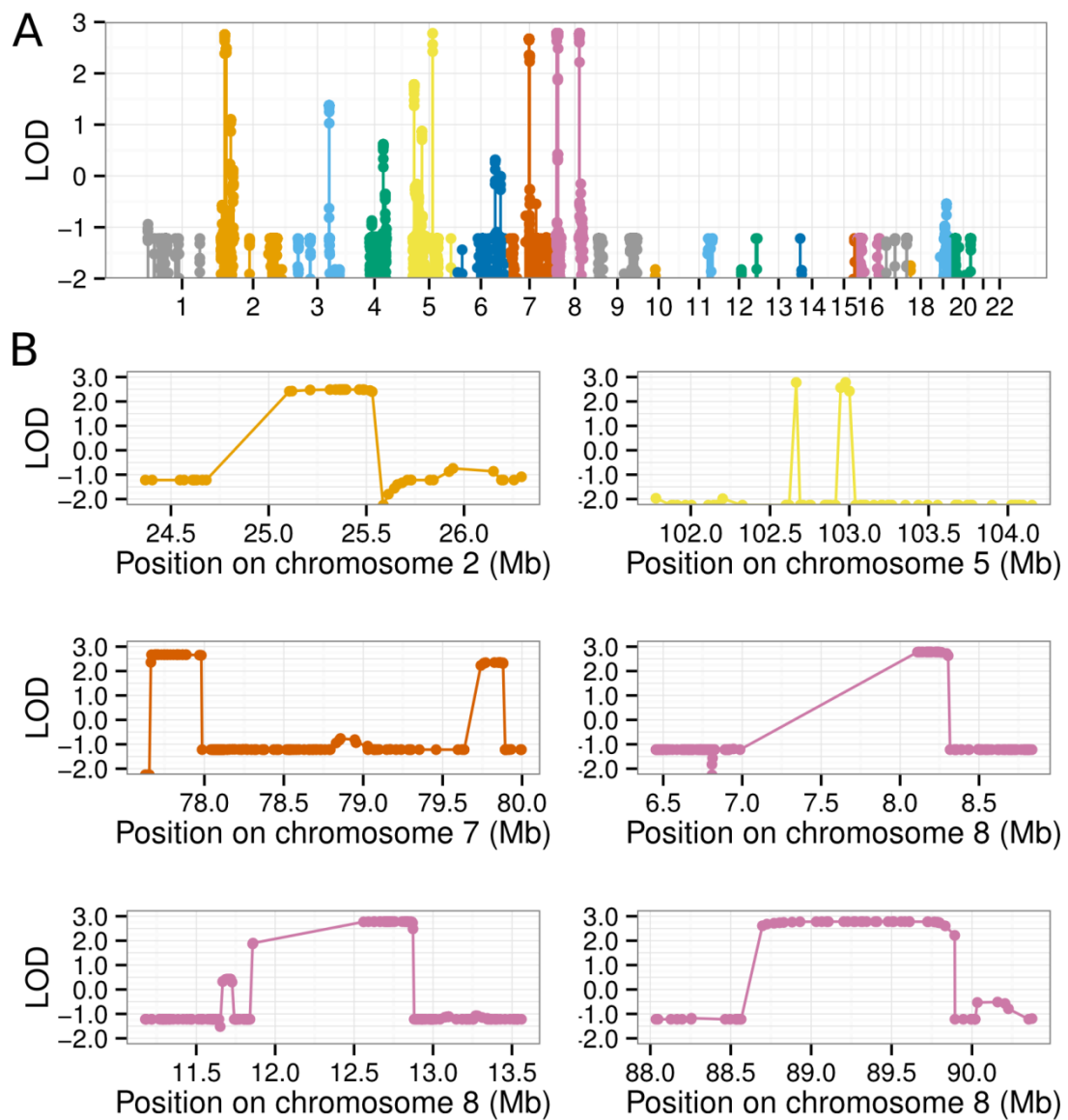
### Cas clinique

Le patient index, P7, est née en 2000 de parents consanguins (Figure 21A). La dissémination exceptionnelle de *M. ulcerans* chez cette patiente est résumée en Figure 22. A 5 ans, elle s'est présentée au CDTUB de Pobè, Bénin, avec une altération de l'état général, une cachexie, de la fièvre et un membre inférieur droit oedématié (Figure 21B). *M. ulcerans* a été identifié dans le liquide de ponction articulaire du genou droit par coloration de Ziehl-Neelsen et par amplification PCR. Une seconde localisation oedémateuse au pied gauche était également diagnostiquée. Le diagnostic d'ostéomyélite du tibia droit, d'arthrite septique du genou droit et d'ostéomyélite du cuboïde gauche étaient confirmés par radiographie. Malgré un traitement prolongé par rifampicine et streptomycine (90 jours) et plusieurs interventions chirurgicales, l'infection a poursuivi sa dissémination. Deux mois après son admission, une lésion oedémateuse du bras droit a révélé une ostéo-arthrite du coude droit, confirmée par l'aspiration d'une substance caséiforme à la ponction, par la coloration de Ziehl-Neelsen du prélèvement et par l'amplification PCR de l'ADN de la bactérie. La fièvre n'a cédé que quelques jours après l'amputation de sa jambe droite, à mi-cuisse, réalisée trois mois après l'admission. A six mois d'évolution, une ostéomyélite du radius gauche était diagnostiquée. A ce stade, les quatre membres étaient donc atteints. A un an d'évolution, la fibula gauche était atteinte.

La patiente a subi de nombreuses interventions chirurgicales de chaque membre lors des 18 mois suivant le diagnostic. Des rechutes ont été diagnostiquées deux, trois et cinq ans après la guérison apparente de l'épisode initial, avec respectivement une ostéomyélite du l'humérus droit, une collection péri-tendineuse du tendon d'Achille gauche et une ostéomyélite du tibia gauche. Elle n'a pas souffert d'autres rechutes depuis (quatre ans sans rechute en 2014). *M. ulcerans* a été identifié de nombreuses fois par Ziehl-Neelsen et PCR dans différents prélèvements de différentes

lésions à différents temps (Figure 22). En dehors de l'UB, les antécédents médicaux de P7 incluent également plusieurs accès palustres standards et une hépatite B.

La sœur de P7, P1, est née en 1995. A 13 ans, elle a développé une large lésion oedémateuse regardant l'ensemble du bras droit (Figure 21D). La PCR et la culture étaient positives pour *M. ulcerans*. Les deux patientes étaient séronégatives pour le VIH et avaient été vaccinées par le BCG à la naissance. Elles ne sont pas atteintes de drépanocytose. Leurs Numération-Formule Sanguine étaient normales. Les deux parents et sept autres frères et sœurs n'ont pas développé d'UB.



**Figure 23. Résultats de l'analyse de liaison par cartographie d'homozygotie**

Analyse prenant en compte atteints et non-atteints

(A) Vue globale de l'ensemble du génome, chromosome 1 à 22

(B) Détail des régions de liaison atteignant un LOD score maximal de 2.8

## Investigations génétiques

Tous les échantillons et 310993 SNPs ont passé le contrôle qualité. Afin de prendre en compte correctement le déséquilibre de liaison entre ces marqueurs, ils ont été groupés en 126704 groupes (clusters) pour analyse de liaison génétique par cartographie d'homozygotie. L'information génétique moyenne pangénomique, calculée par MERLIN, était de 0.72. Un LOD score maximum théorique peut être calculé pour une configuration familiale donnée, un marqueur génétique complètement informatif et un modèle génétique spécifié. Dans notre cas (2 enfants atteints, 3 enfants non-atteints, consanguinité des parents de type cousins au second degré et un modèle génétique récessif à pénétrance complète), ce LOD score maximum vaut 2.8, ce qui correspond à une pvalue minimale de  $3.10^{-3}$ . Sept régions ont atteint ou fortement approché ce LOD score maximum théorique et étaient situées sur les chromosomes 2, 5, 7 et 8 (Figure 23), sur une étendue totale de 5.4 mégabases. La région du chromosome 2 et la deuxième région du chromosome 7 approchaient en fait le LOD score maximal, avec respectivement des LOD scores à 2.3 et 2.5, pondérés par une information à 0.73 et 0.75. Les autres régions avaient un LOD score maximal supérieur à 2.7 et une information supérieure à 0.85. La liste complète des gènes dans ces régions (Ensembl v75) est donnée en Tableau 5. De manière intéressante, deux régions de liaison sur le chromosome 8 contiennent un cluster de gènes de codant pour les bêta-défensines.

Dans l'optique d'un modèle récessif, l'analyse de l'exome a permis l'identification de 5397 variations homozygotes exoniques partagées par les deux patientes. Après élimination des variations connues listées dans les bases de données publiques (dbSNP, Exome Variant Server), 9 variations communes aux deux patientes étaient identifiées. Aucune n'était localisée dans les régions de liaison et aucune ne concernait de gène potentiellement candidat. La qualité de l'alignement était faible dans la région des défensines et a empêché une identification de variants dans cette région.



Tableau 5. Liste des gènes localisés sous les pics de liaison (1/2)

Chr2	Description	Gene Biotype
NCOA1	nuclear receptor coactivator 1	protein_coding
PTRHD1	peptidyl-tRNA hydrolase domain containing 1	protein_coding
CENPO	centromere protein O	protein_coding
ADCY3	adenylate cyclase 3	protein_coding
DNAJC27	DnaJ (Hsp40) homolog, subfamily C, member 27	protein_coding
EFR3B	EFR3 homolog B (S. cerevisiae)	protein_coding
RP11-509E16.1		lincRNA
POMC	proopiomelanocortin	protein_coding
RP11-458N5.1		lincRNA
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	protein_coding
MIR1301	microRNA 1301	miRNA

Chr7	Description	Gene Biotype
MAGI2	membrane associated guanylate kinase	protein_coding
GNAI1	guanine nucleotide binding protein (G protein), alpha inhibiting	protein_coding
AC003988.1		lincRNA

Chr8.1	Description	Gene Biotype
AF228730.2		miRNA
FAM66B	family with sequence similarity 66, member B	lincRNA
ZNF705G	zinc finger protein 705G	protein_coding
DEFB4B	defensin, beta 4B	protein_coding
DEFB103B	defensin, beta 103B	protein_coding
SPAG11B	sperm associated antigen 11B	protein_coding
AC130360.1		miRNA
DEFB104B	defensin, beta 104B	protein_coding
DEFB106B	defensin, beta 106B	protein_coding
DEFB105B	defensin, beta 105B	protein_coding
DEFB107B	defensin, beta 107B	protein_coding
PRR23D1	proline rich 23 domain containing 1	protein_coding
PRR23D2	proline rich 23 domain containing 2	protein_coding
DEFB107A	defensin, beta 107A	protein_coding
DEFB105A	defensin, beta 105A	protein_coding
DEFB106A	defensin, beta 106A	protein_coding
DEFB104A	defensin, beta 104A	protein_coding
SPAG11A	sperm associated antigen 11A	protein_coding
AC130365.1		miRNA
DEFB103A	defensin, beta 103A	protein_coding
DEFB4A	defensin, beta 4A	protein_coding
ZNF705B	zinc finger protein 705B	protein_coding
FAM66E	family with sequence similarity 66, member E	lincRNA
MIR548I3	microRNA 548i-3	miRNA

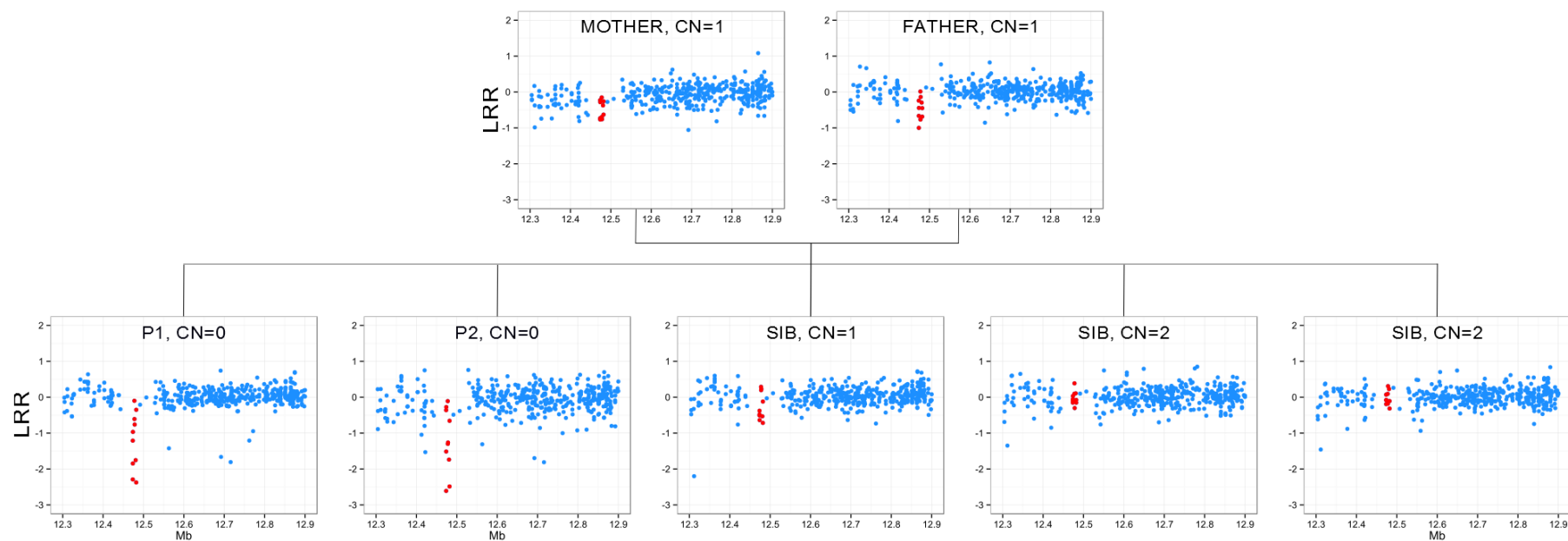
**Tableau 6. Liste des gènes localisés sous les pics de liaison (2/2)**

<b>Chr8.1 (suite)</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>AC068020.1</b>		miRNA
<b>LRLE1</b>	Liver-related low express protein 1	protein_coding
<b>SGK223</b>	Tyrosine-protein kinase Sgk223	protein_coding
<b>CTA-398F10.1</b>		lincRNA
<b>CTA-398F10.2</b>		lincRNA

<b>Chr8.2</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>DEFB135</b>	defensin, beta 135	protein_coding
<b>DEFB134</b>	defensin, beta 134	protein_coding
<b>RP11-481A20.11</b>	Protein LOC101060662	protein_coding
<b>DEFB130</b>	beta-defensin 130-like precursor	protein_coding
<b>ZNF705D</b>	zinc finger protein 705D	protein_coding
<b>USP17L2</b>	ubiquitin specific peptidase 17-like family member 2	protein_coding
<b>AC145124.1</b>		miRNA
<b>FAM86B1</b>	family with sequence similarity 86, member B1	protein_coding
<b>DEFB130</b>	defensin, beta 130	protein_coding
<b>LINC00965</b>	long intergenic non-protein coding RNA 965	lincRNA
<b>AC087203.1</b>		miRNA
<b>FAM86B2</b>	family with sequence similarity 86, member B2	protein_coding
<b>RP11-419I17.1</b>		lincRNA
<b>AC130352.1</b>		miRNA
<b>AC068587.2</b>		miRNA
<b>AC068587.1</b>		miRNA
<b>MIR5692A1</b>	microRNA 5692a-1	miRNA
<b>LONRF1</b>	LON peptidase N-terminal domain and ring finger 1	protein_coding
<b>MIR3926-2</b>	microRNA 3926-2	miRNA
<b>RP11-252C15.1</b>		lincRNA
<b>LINC00681</b>	long intergenic non-protein coding RNA 681	lincRNA
<b>AC090133.1</b>		miRNA
<b>KIAA1456</b>	KIAA1456	protein_coding

<b>Chr8.3</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>CNBD1</b>	cyclic nucleotide binding domain containing 1	protein_coding
<b>DCAF4L2</b>	DDB1 and CUL4 associated factor 4-like 2	protein_coding
<b>MMP16</b>	matrix metalloproteinase 16 (membrane-inserted)	protein_coding
<b>RP11-642C5.1</b>		lincRNA

Seul les gènes codant pour une protéine, un ARN non-codant de type miRNA ou lincRNA sont listés ici. Les intervalles exacts des régions interrogées sur Ensemblv75 (GRCh37) sont : 2:24680948-25585907, 5:102620674-102689268, 5:102913885-103038495, 7:77653409-77985583, 7:79635822-79894010, 8:6986241-8316581, 8:11840071-12880628, 8:88564546-89894046. Aucun gène n'est annoté dans les régions d'intérêt du chromosome 5.



**Figure 24. Résultats de l'analyse des variations en nombre de copies - chromosome 8**

Chaque cercle bleu représente la valeur de l'intensité du signal à un marqueur du génome, telle que mesurée par le Log R Ratio (LRR). CN: nombre de copies. Les parents et un enfant non-atteint sont hétérozygotes pour la délétion, deux enfants non atteints ont un nombre normal de copies, et les deux enfants atteints ont une délétion complète d'une région de 8.5 kb.

L'analyse des variations en nombre de copies a été effectuée sur l'ensemble du génome. L'ensemble des SNPs a été conservé, puisque dans cette approche le déséquilibre de liaison n'est pas un obstacle et que tous les SNPs, même monomorphes, sont informatifs. De fait, la puce Affymetrix 6.0 inclut également 946 000 marqueurs additionnels monomorphes dédiés à la détection de variations en nombre de copies, soit un total d'environ 1 800 000 marqueurs pour cette analyse. Nous avons identifié une seule délétion présente à l'état homozygote chez les deux atteints, hétérozygote chez les parents, et hétérozygote ou absente chez les enfants non-atteints. Cette délétion de 8.5 kb est localisée au centre de la région de liaison du chromosome 8. Graphiquement, elle peut être appréciée sur la Figure 24, qui montre le log R ratio (LRR), une mesure de l'intensité du signal à chaque SNP, et qui est très généralement proportionnelle au nombre de copies. Une valeur nulle signifie un nombre de copies normal de 2, une valeur négative un nombre de copies inférieur à 2. La délétion ségrège parfaitement avec le phénotype selon un mode de transmission récessif. Dans l'ensemble du génome, aucun autre CNV détecté par l'analyse n'avait cette caractéristique.

## Génétique mendélienne de l'UB : discussion

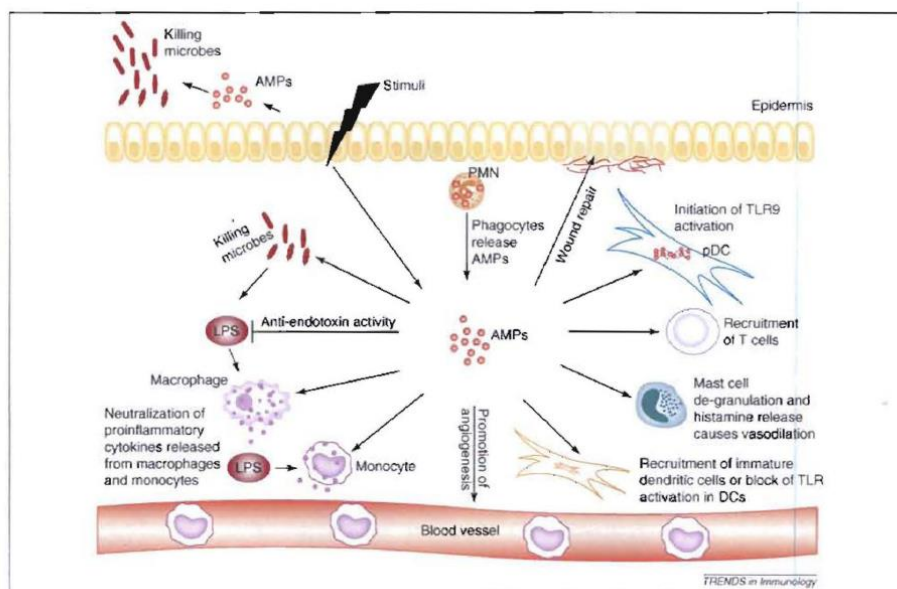
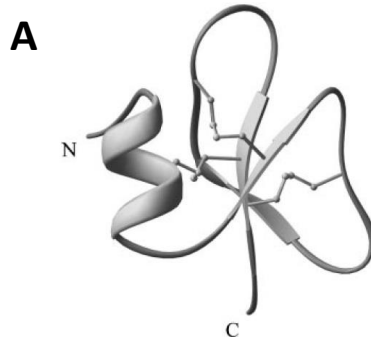
Nous rapportons le phénotype clinique et l'exploration génétique d'une famille consanguine présentant une forme extrême d'UB avec une ségrégation mendélienne. Le patient index a présenté un UB multifocal extrêmement agressif avec de nombreuses atteintes osseuses et articulaires ayant abouti à l'amputation de sa jambe droite. Sa sœur a présenté un UB sévère du bras droit. Nous avons identifié une délétion homozygote partagée par les deux patients au sein d'une région de liaison sur le chromosome 8. Cette exploration génétique a donc généré une hypothèse de recherche inattendue : les défensines pourraient jouer un rôle dans le contrôle de l'infection par *M. ulcerans*.

Les défensines font partie de la famille des peptides antimicrobiens de l'immunité innée, peptides exprimés dans les épithéliums et muqueuses de la peau, des poumons, de l'arbre urinaire, et de l'intestin, à l'interface du corps et de l'environnement extérieur (193,194). Les neutrophiles, monocytes, macrophages et cellules dendritiques sont également producteurs de défensines. Elles sont nombreuses, et pour certaines d'entre elles dupliquées de manière spécifique chez l'homme (paralogues). Le cluster principal des défensines est situé en 8p23, mais la séquence du génome humain a révélé *in silico* d'autres défensines probables, sur les chromosomes 6 et 20 essentiellement mais aussi 4 et 11, pour un total d'environ 38 gènes et 11 pseudogènes prédits (Ensembl 75). Toutes les défensines ont en commun six résidus cystéine formant trois ponts disulfures, fortement conservés chez les mammifères (Figure 25A).

Les  $\beta$ -défensines sont un candidat séduisant pour expliquer les formes extrêmes de la maladie infectieuse ulcérate de peau qu'est l'UB (Figure 25B). D'une part, *DEFB4* et *DEFB103*, produisant hBD-2 et hBD-3, sont deux défensines inductibles dans la peau, produites par les kératinocytes et les cellules phagocytaires après stimulation par le  $\text{TNF}\alpha$ , l'IL1- $\beta$  et l'IFN $\gamma$ . Cette production peut être le fait de deux stimuli non concomitants et tous deux pertinents dans l'UB : la présence d'un microbe ou l'effraction traumatique du tissu cutané (193–195). Elles ont une activité antimicrobienne prouvée, *in vitro*, contre *Staphylococcus aureus* et *Pseudomonas aeruginosa*. La

souris invalidée dans le gène *defb1* présente une élimination pulmonaire retardée de *Haemophilus influenzae* et une colonisation vésicale par *Staphylococcus aureus*. Cette activité antimicrobienne est médiée par une déstabilisation de la membrane bactérienne. Une équipe explore spécifiquement le rôle des défensines dans le contrôle de l'infection osseuse et montre en particulier leur production massive par les ostéoblastes (196–198). Elles ont une activité proinflammatoire majeure, déclenchant la production par les kératinocytes de nombreuses cytokines dont IL-6, IL-10 et RANTES, et démontrant un effet chimiotactique sur les cellules dendritiques immatures et les lymphocytes T mémoire CD4+CD45RO+. Il est intéressant de noter qu'après une lésion cutanée traumatique, en l'absence d'infection, les défensines murines mBD-1, mBD-2 et mBD-14 (protéines produites par les orthologues de *DEFB1*, *DEFB4* et *DEFB103*) sont produites en grande quantité dans la peau (195). Elles ont un rôle déterminant dans la cicatrisation cutanée, en promouvant l'angiogénèse et la réépithélisation (193).

Le rôle des défensines dans les maladies mycobactériennes est en cours d'exploration (199). L'infection de cellules épithéliales humaines par le BCG ou par *M. tuberculosis* induit une production de la défensine hBD-2 par le gène *DEFB4*. Les monocytes humains surexpriment les  $\beta$ -défensines après stimulation par CD40L et l'IFN $\gamma$  (200). La surexpression de *DEFB4* par des macrophages humains divise par 10 la multiplication de *M. tuberculosis* (201). *In vivo*, chez la souris, l'injection intratrachéale de *M. tuberculosis* induit une production majeure de mBD-3 et mBD-4 (deux orthologues de *DEFB4*) pendant le premier mois de l'infection pulmonaire (202). Dans une approche vaccinale chez la souris, l'adjonction de  $\beta$ -défensine à une vaccination par le BCG améliore la survie après infection expérimentale par *M. tuberculosis* (203). Les défensines sont également surexprimées dans les réactions de réversion de la lèpre (204). Dans le cadre de l'UB, il a été rapporté qu'une souche béninoise de *M. ulcerans* induisait une forte production des défensines mBD-3 et mBD-4 identifiée par RT-PCR et par immunohistochimie, et ce au moins pendant le premier mois de l'infection (205).

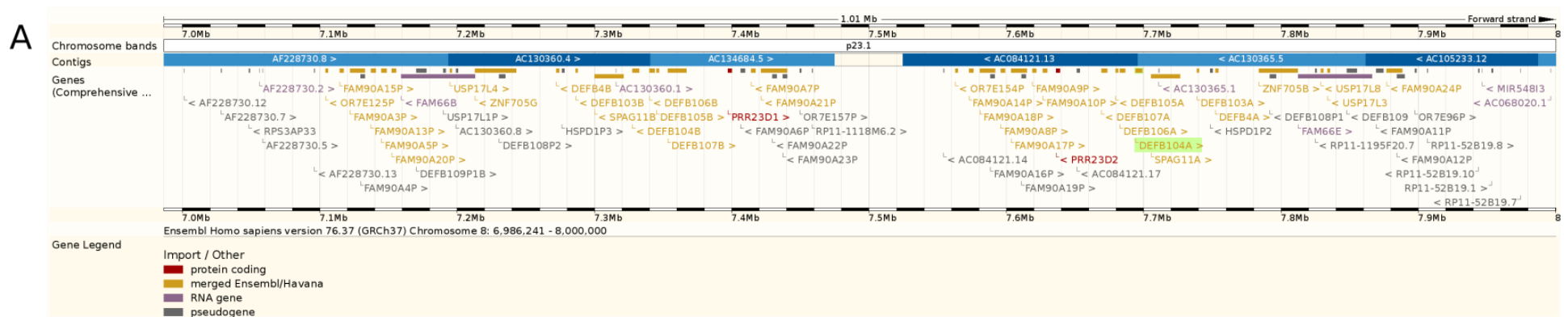


**Figure 25. Structure et fonction des défenses**

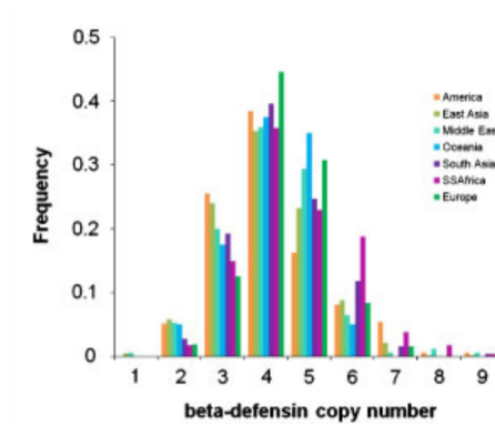
(A) Toutes les défenses, comme ici hBD2, sont caractérisées par la présence de trois feuillets  $\beta$  stabilisés par trois ponts disulfures.

(B) Schéma récapitulatif des différentes fonctions attribuées aux peptides antimicrobiens. Les  $\beta$ -défenses ont été tout particulièrement impliquées dans la destruction microbienne et la cicatrisation cutanée.

*Adapté de (193,206)*



**B**



**Figure 26. Le locus des défensines en 8p23**

(A) Cluster des défensines sur Ensembl 75. On identifie les différentes défensines (en jaune; *DEFB4* est surlignée en vert), ainsi que le trou dans les contigs de la séquence de référence (ligne bleue)

(B) Distribution du nombre de cluster de défensines par individu dans la monde en 8p23



Il existe certaines limites à l'hypothèse d'une implication des défensines dans la protection contre *M. ulcerans*. La région du cluster des défensines est particulièrement complexe (Figure 26). Dans la population générale, il existe une duplication à grande échelle du cluster des défensines en 8p23, conduisant à un nombre normal de copies compris entre deux et sept par génome diploïde (parfois neuf à douze, variation détectable macroscopiquement et connue sous le nom de variants euchromatiques du 8p23) dans les populations du panel du Centre d'Etude du Polymorphisme Humain (CEPH) (Figure 26B) (207,208). Le nombre de copies le plus fréquent est de quatre (deux par chromosome 8) et correspond à celui représenté par la séquence de référence du génome humain telle que consultable sur Ensembl, qui propose deux copies inversées autour de la position 7 Mb (Figure 26A). Une copie correspond à une séquence d'environ 250kb qui inclut les gènes *DEFB4*, *DEFB103*, *DEFB104*, *DEFB105*, *DEFB106*, *DEFB107*, mais pas *DEFB1*. L'assemblage de cette région est particulièrement difficile, et il persiste d'ailleurs un trou dans la carte physique de référence du génome humain entre ces deux copies (209). Un deuxième cluster indépendant de défensines (défensines *DEFB130*, *DEFB134*, *DEFB135*, *DEFB136*), distinct et non dupliqué, se trouve environ à 4.5 mégabases en aval et correspond à notre deuxième région de liaison sur le chromosome 8.

L'analyse des variations en nombre de copies (CNV) par la méthode PennCNV-joint donne un nombre de copies de 2 pour le cluster des défensines *DEFB4* à *DEFB107*. La méthode utilisée ici pour l'estimation du nombre de copies n'est pas optimale pour des régions si complexes du génome humain, en particulier parce que le nombre de copies moyen dans la population est élevé (>4). Une hypothèse intrinsèque aux méthodes de détection de CNV par puces de génotypage est que le nombre de copies présent dans les échantillons de référence est de 2, ce qui n'est probablement pas le cas ici. PennCNV ne prévoit d'ailleurs que 6 catégories : 0, 1, 2 (avec ou sans perte d'hétérozygotie), 3 et 4 copies d'une séquence donnée. La disparité du nombre de copies "normales" incite à la prudence quant à l'évaluation du nombre de copies dans nos échantillons d'intérêt à ce locus (189,210). Les résultats contradictoires publiés sur la comparaison du nombre de copies du cluster des défensines dans des enquêtes cas-témoin de la maladie de Crohn ont été attribués à la

faible précision de leur estimation (211,212). On peut cependant dire qu'il ne semble pas y avoir de différence majeure avec le nombre moyen de copies dans les 200 échantillons de référence internes au laboratoire utilisés pour l'analyse de la famille BU-A.

L'évaluation du nombre exact de copies du cluster des défensines semble d'une importance cruciale pour incriminer l'une ou plusieurs d'entre elles dans la pathogénèse. En effet, la question majeure et non résolue est celle de la redondance des différentes défensines. Il est à noter que l'expression des différentes défensines n'est pas toujours mesurée. La base de données bioGPS comporte ainsi des données pour *DEFB4*, mais pas *DEFB103-DEFB107* ni *DEFB130-DEFB136* (213). En supposant qu'elles ne soient pas redondantes, leur prêter une pathogénicité alors que le locus est habituellement répété suppose qu'un ou plusieurs mécanismes soient vérifiés : une mutation homozygote d'une défensine s'est propagée à son ou ses paralogues (conversion génique, recombinaison non homologue); plusieurs événements mutationnels ont touché les différents paralogues, de telle manière que l'ensemble des copies sont inactivées; l'ensemble du locus est manquant chez les deux atteintes; l'ensemble du locus est réprimé chez les deux atteintes. La délétion reste une hypothèse séduisante, puisque le typage du cluster des défensines chez plus de 1500 individus n'a jamais mis en évidence d'individus avec un nombre de copies nul, soulignant l'importance fonctionnelle probable de ce locus (214). Enfin, la répression pourrait se faire à travers la modification d'un amplificateur ("enhancer") ou d'un ARN non-codant régulateur (215). Le criblage de bases de données n'a pas encore identifié de tels motifs régulateurs au niveau du cluster *DEFB4-DEFB107*.

En revanche, l'analyse de CNV a identifié une délétion homozygote de 8.5 kb dans une deuxième région du chromosome 8 (aux environs de la position 12Mb) ségrégeant parfaitement avec le phénotype dans la famille explorée. Comme le nombre de copies de ce locus dans la population générale semble être de deux (à l'inverse de la région de liaison précédente) l'analyse par PennCNV peut être considérée comme fiable. Il est néanmoins connu que les algorithmes de détection de CNV

ont une résolution modeste pour estimer les bornes d'un CNV. Une sous-estimation de la taille des CNV est fréquente, ce qui implique que la délétion identifiée pourrait être plus large que 10kb (216). De manière intéressante, deux ARNs courts de type microRNA ont été identifiés à moins de 40 kilobases de la délétion. Au début de ma thèse, la version 62 d'Ensembl rapportait également la présence d'un ARN intergénique non-codant de grande taille (lincRNA - long intergenic non coding RNA) nommé *AC068587.3* (ENSG00000247397), exactement à la localisation de la délétion. Cet ARN, prédit par Ensembl 62 a néanmoins disparu de la dernière version des annotations (Ensembl v75). Le potentiel régulateur des ARN non-codants laisse à penser que l'expression des  $\beta$ -défensines pourrait être compromise chez les deux patientes. Ce potentiel régulateur a été particulièrement documenté dans la réponse immunitaire (217). Néanmoins, la complexité globale de ce locus a empêché jusqu'ici une validation fonctionnelle définitive de l'implication des défensines dans le phénotype clinique de nos patientes. Le séquençage de l'exome n'a pas fourni d'autres candidats d'intérêt. En ce qui concerne l'hypothèse des défensines, il est intéressant de noter que cette technologie de pointe est également tenue en échec par la complexité de cette région (218). Pour le moment, l'identification d'une seconde famille mendélienne partageant la même région de liaison que celle-ci semble être le chemin le plus vraisemblable pour conforter notre hypothèse causale.





# Genétique humaine de l'ulcère de Buruli : aspects complexes

---



## Collection des données cliniques et biologiques

### Etude de faisabilité

#### Etude de faisabilité globale : voyages de terrain

Quelques mois avant mon inscription en thèse, la faisabilité du projet « Génétique Humaine de l'Ulcère de Buruli » était encore incertaine, malgré l'obtention de financements adéquats (Fondation pour la Recherche Médicale, puis Agence Nationale de la Recherche). Le lien avec le CDTUB de Pobè reposait entièrement sur la collaboration naissante avec l'équipe Biologie et Ecologie de *M. ulcerans* dirigée par Laurent Marsollier au Centre Hospitalo-Universitaire d'Angers. Par la suite, le projet bénéficiera du soutien de l'association Raoul Follereau France et du gouvernement béninois.

Sur le plan pratique, il restait à déterminer si (i) le flux de malades au CDTUB de Pobè allait être suffisant pour recruter le nombre de patients nécessaire pour atteindre une puissance statistique adéquate, (ii) nous allions recueillir un échantillon cas-témoin ou familial, (iii) le contexte tropical d'une région reculée du Bénin allait permettre la constitution d'une large banque biologique, (iv) le projet pourrait se dérouler dans des conditions éthiques satisfaisantes. Un premier déplacement sur le terrain en juin 2010, puis un second en juin 2011, ont permis de lever plusieurs de ces interrogations.

Le fonctionnement du CDTUB de Pobè a été présenté précédemment dans ce manuscrit. Rappelons néanmoins quelques éléments ici. Environ cinquante personnes travaillent au CDTUB, dont trois médecins et un chirurgien à temps partiel. La capacité d'hospitalisation est de 58 lits. Le laboratoire d'analyses biologiques permet la réalisation des examens essentiels. Les radiologies sont effectuées à l'Hôpital de zone de Pobè. Le CDTUB coordonne les soins de l'UB en hospitalisation (Pobè) et en ambulatoire (Pobè et environ 15 dispensaires/ postes avancés de soins au cœur des villages) dans un rayon d'environ 30 kilomètres.



L'organisation des soins prévoit trois tournées hebdomadaires dans les villages endémiques. Tous les patients sont revus à distance du traitement pour une détection des rechutes et une évaluation des séquelles fonctionnelles. Il apparaît que le lien entre l'équipe du CDTUB et les patients et leur famille est durablement établi. Le nombre annuel de patients traités est de l'ordre de 200. Une partie de leurs caractéristiques cliniques est informatisée sous format Excel depuis 2003. La grande majorité des anciens patients semble toujours habiter la région. Ces différentes observations nous permettent d'envisager un recrutement mixte rétrospectif et prospectif.

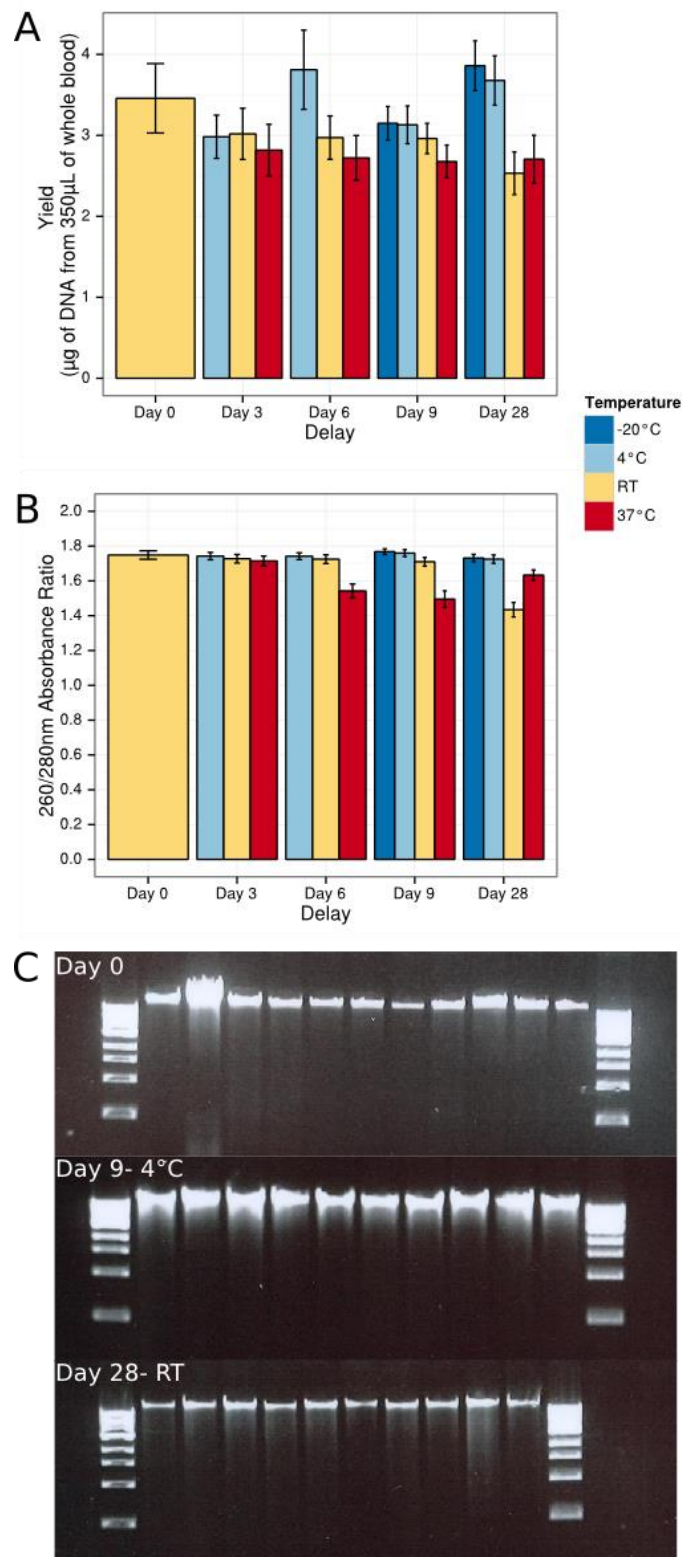
La population de cette région du Bénin est homogène, avec une population Fon et Yoruba. Ibadan, la ville nigériane où ont été recrutés les 180 individus Yoruba du panel génétique international HapMap est à 250 kilomètres. L'immense majorité parle fon, certains parlent français, et rares sont ceux appartenant à des populations isolées avec leur propre dialecte et culturellement endogames. Le taux d'alphabétisation semble modeste. La structure des familles est extrêmement difficile à reconstituer. A titre d'exemple, l'ordre de naissance dans la fratrie n'est pas sans équivoque. Le nom de famille d'usage n'est pas nécessairement unique au sein d'une fratrie. L'homogénéité génétique et l'accès difficile à la généalogie nous incitent à opter pour une étude de type cas-témoin.

La balance entre le bénéfice à long terme de notre programme de recherche pour la compréhension et la prise en charge de la maladie et la contribution individuelle des patients (consentement à une étude génétique, prélèvement sanguin unique) nous semble favorable sur un plan éthique. Sur place, la qualité du lien entre l'équipe du CDTUB et ses patients et la présence systématique d'un traducteur nous donnent la conviction que l'étude pourra se dérouler dans des conditions d'information et de consentement respectueuses de la personne. A l'échelle du CDTUB, il semble également que la collaboration à ce projet de recherche soit bénéfique. Il en résulte un transfert bilatéral de connaissance entre les équipes françaises et béninoises. L'équipe du CDTUB partage son expertise clinique quotidienne de l'UB ; l'équipe de Laurent Marsollier organise un

transfert scientifique de connaissance pour la réalisation progressive sur place du diagnostic moléculaire d'UB par PCR (effective en 2014) ; l'équipe d'Alexandre Alcaïs apporte son expertise scientifique en génétique et santé publique. Le projet a été approuvé par les comités éthiques du Bénin (IRB00006860) et du Centre Hospitalo-Universitaire d'Angers.

En revanche, la question de la gestion pratique des échantillons biologiques n'est pas résolue à ce stade. En effet, les villages endémiques autour de Pobè appartiennent soit au bassin du fleuve Ouémé, zone largement inondée lors de la saison humide, soit au plateau qui domine la rive Est de ce fleuve. Le climat est tropical et comporte une grande saison sèche (novembre à mars), une petite saison sèche (août) et deux saisons des pluies (avril à juillet, septembre à octobre). La température moyenne est relativement constante dans l'année autour de 28°C à l'ombre. L'accès aux villages et dispensaires périphériques dépend de pistes de qualité inégale. Leur parcours requiert un véhicule tout-terrain et l'expérience d'un chauffeur professionnel, en particulier en saison des pluies. Autrement dit, les prélèvements sanguins faits en périphérie devraient circuler plusieurs heures dans des conditions chaudes et chaotiques avant de rejoindre le CDTUB.

Les capacités logistiques sur place au début de l'étude rendent impossible l'extraction locale des ADN. Ceux-ci doivent donc être envoyés en France par une compagnie postale privée dont les bureaux sont à Cotonou, à trois heures de route. Il n'y a aucune possibilité d'envoi congelé ou frais. Par ailleurs, il serait difficile de faire plus d'un envoi par semaine. Les colis sont livrés en 2 à 4 jours. La question posée est donc la suivante : est-il possible d'obtenir un ADN de quantité et qualité suffisante pour des applications génétiques de pointe, pangénomiques, à partir de sang prélevé 4 à 10 jours auparavant et transporté entre 0 et 40°C par jeep et avion ?



**Figure 27. Impact du délai d'extraction et de la température de conservation sur le rendement et l'intégrité de l'ADN : conditions de laboratoire**

(A) Rendements quantitatifs moyens en µg d'ADN extrait à partir de 350µL de sang total (n=20):

(B) Rapports d'absorbance moyens (n=20)

(C) Migration sur gel d'agarose1% de 100 à 150 ng d'ADN (n=10, représentatifs)

*Les barres d'erreur représentent l'erreur type de la moyenne.*

## Etude de faisabilité logistique : du prélèvement sanguin au génotypage

De multiples facteurs sont impliqués dans le succès d'une étude d'association pangénomique : la validité de l'hypothèse génétique pour le trait d'intérêt, la qualité et l'homogénéité du phénotype étudié, une taille d'échantillon conférant une puissance statistique suffisante et enfin, la qualité du génotypage effectué. Le recueil de plusieurs microgrammes d'ADN de haute qualité est nécessaire pour l'interrogation fiable de millions de marqueurs génétiques sur le génome. En 1993, une étude vétérinaire a montré que le sang bovin pouvait être conservé à 4°C pendant un mois. En effet, les rendements diminuaient raisonnablement pour un stockage à 4°C mais largement pour un stockage à 23 ou 37°C (219). Plus récemment, il a été montré que la quantité et qualité de l'ADN extrait de sang total humain après 24 heures à 4°C était stable, mais qu'elle tendait à diminuer après 7 jours (220,221). Les protocoles typiques d'extraction d'ADN recommandent que l'extraction d'ADN depuis sang total soit faite dans les 24 heures qui suivent la prise de sang (par exemple, kit BACC2, GE Healthcare). Cette recommandation peut néanmoins être impossible à suivre sur le terrain. Dans notre cas, un délai de 4 à 10 jours et de fortes variations thermiques étaient à prévoir. Avant de lancer une étude impliquant le prélèvement et le génotypage de plus de mille individus, nous avons évalué la quantité et qualité de l'ADN extrait de sang total conservé dans différentes conditions jusqu'à un mois.

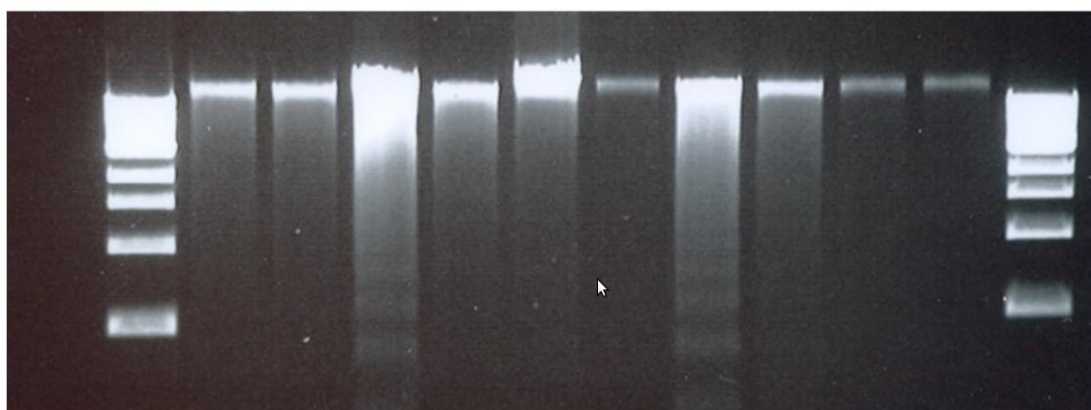
Nous avons d'abord réalisé des tests en conditions contrôlées au laboratoire. Six millilitres de sang total ont été collectés chez 20 donneurs sains (Paris, France) en tube plastique K<sub>2</sub>EDTA (Beckton-Dickinson). Chaque échantillon était réparti immédiatement après le prélèvement en aliquots de 350µL stockés à -20, +4, +22 ou +37°C pendant 3, 6, 9 ou 28 jours. L'ADN était extrait immédiatement d'un aliquot. L'extraction d'ADN était réalisée par l'automate de purification iPrep™ (PureLink™ gDNA Blood Kit, Invitrogen). Tous les ADNs ont ensuite été dosés par la méthode Quant-iT™ PicoGreen® (Invitrogen), méthode de référence ultrasensible et spécifique pour le dosage de l'ADN double-brin. La qualité et la pureté de l'ADN ont été évaluées par la mesure du ratio des

absorbances mesurées à 260 et 280 nanomètres au spectrophotomètre et par électrophorèse de 150 nanogrammes d'ADN sur gel d'agarose 1% (Tris-Borate-EDTA). Un échantillon pilote a été génotypé sur puce Illumina 650K (Centre National de Génotypage, Evry, France).

Sur l'ensemble des délais et températures de conservation, les rendements moyens en ADN pour 350µL de sang total étaient compris entre 2.5 et 3.8 µg (Figure 27A). Le stockage pendant un mois n'a pas modifié significativement les rendements, observation en contraste marqué avec l'étude de Cushwa et Medrano (219). Les rapports d'absorbance 260/280 moyens étaient élevés (de l'ordre de 1.7, Figure 27B) et constants jusqu'à J28 à -20°C et 4°C, J9 à 23°C et J3 à 37°C. Ces rapports tombaient à 1.44 à J28 à 23°C et 1.5 dès J6 à 37°C. Les migrations sur gel révèlent en revanche que l'intégrité de l'ADN n'était pas compromise par la durée ou la température de stockage (Figure 27C). De façon notable, un échantillon conservé à 4°C pendant 28 jours a été génotypé sur puce pangénomique avec un taux de génotypage de 99.7% malgré une dégradation marquée.

Ces données préliminaires nous ont permis d'envisager une deuxième phase de test concernant le sang total prélevé au Bénin et extrait en France après une dizaine de jours de trajet à température variable et dans des conditions de stockage variables. Deux échantillons de 3mL de sang total en tube EDTA ont été prélevés chez 5 individus vivant en périphérie du CDTUB de Pobè, Bénin. Après prélèvement, les échantillons ont été transportés par jeep à température ambiante locale (30-35°C) pendant 2 à 4 heures puis stockés à 4°C (réfrigérateur) ou 20-25°C (pièce climatisée) pour 3 à 4 jours. Ils ont ensuite été transportés à l'aéroport international à trois heures de route, confiés à température ambiante à l'entreprise privée DHL pour son envoi hebdomadaire (avion de marchandise s'arrêtant dans plusieurs capitales africaines avant de rejoindre Paris ; trajet de 2-3 jours ; >5000 kilomètres ; température en soute ~10°C). L'extraction d'ADN n'utilisait plus la méthode iPrep, incompatible avec ces volumes de sang, mais le kit BACC2 illustra Nucleon Genomic DNA (GE Healthcare).

	Day 6				Day 9					
	4°	23°	4°	23°	4°	23°	4°	23°	4°	23°
Yield (µg)	12.7	6	5.8	8.1	5.4	29.7	7	10.1	14.7	15.5
A260/A280	1.71	1.8	1.83	1.81	1.69	1.75	1.48	1.64	1.72	1.78



**Figure 28. Impact du délai d'extraction et de la température de conservation sur le rendement et l'intégrité de l'ADN : conditions de terrain**

Le rendement quantitatif en µg d'ADN par mL de sang total, le rapport d'absorbance 260/280nm et le gel de migration sur agarose sont montrés pour 10 échantillons conservés au Bénin à 4°C ou en pièce climatisée avant leur transfert en France par jeep et avion dans un délai de 6 à 9 jours après la prise de sang.

Quatre échantillons ont donc été extraits à J6 après le prélèvement, et six échantillons à J9. Les rendements étaient de 5 à 30 µg par millilitre de sang total (Figure 28). Les rapports moyens d'absorbance 260/280 à 4°C et 23°C étaient de 1.77 et 1.80 à J6. Ils étaient de 1.63 et 1.72 à J9. La migration sur gel a montré une excellente intégrité de l'ADN. Six échantillons ont été génotypés sur puce pangénomique avec un taux de génotypage minimum de 98.2% et moyen de 99.4%. Le taux de génotypage de l'échantillon qui apparaît le plus dégradé (4°C, J6, Figure 28, puit n°4) était de 99.6%. La confirmation finale de la validité de notre processus logistique viendra avec l'extraction d'ADN d'environ 1600 échantillons de sang total en provenance hebdomadaire du Bénin. L'extraction d'ADN a abouti à une quantité insuffisante d'ADN pour environ 50 échantillons, soit environ 3%. Parmi les 816 échantillons de la cohorte primaire, seuls 8 (1%) ont un taux de génotypage inférieur à 95%.

Ces résultats sont d'une importance cruciale pour les équipes sur le terrain dans des zones du monde où (i) les températures ambiantes sont élevées, (ii) les transports difficiles, (iii) l'approvisionnement stable en électricité et donc la chaîne du froid ne sont pas assurés. Nous avons démontré que des échantillons de sang sur EDTA pouvaient être fortement malmenés pendant environ dix jours mais rester de qualité suffisante pour permettre l'application de techniques pangénomiques de pointe. Nous rapportons également ici le séquençage de l'exome de deux échantillons, explorés dans la partie précédente de la thèse. D'après nos tests, réalisés jusqu'à un mois post-prélèvement, la condition optimale de conservation du sang total semble être de -20°C ou 4°C. Il nous semble préférable d'opter pour 4°C afin d'éviter des cycles répétés de congélation/décongélation. Des délais de plus d'un mois seront en général évitables. On trouve dans la littérature la mention d'extraction d'ADN sur tubes de sang conservés à 4°C jusqu'à 1 an ou à -20°C pendant 30 ans, sans toutefois d'application génétique à haut-débit comme ici (220,222). L'immense majorité des études d'association pangénomique a été réalisée sur des populations caucasiennes dans des pays dits développés (223). Notre étude montre que la crainte de limitations techniques liées à la qualité de l'ADN prélevé dans les pays en voie de développement peut être levée afin de promouvoir la recherche biomédicale de pointe dans toutes les régions du monde.

## Recrutement

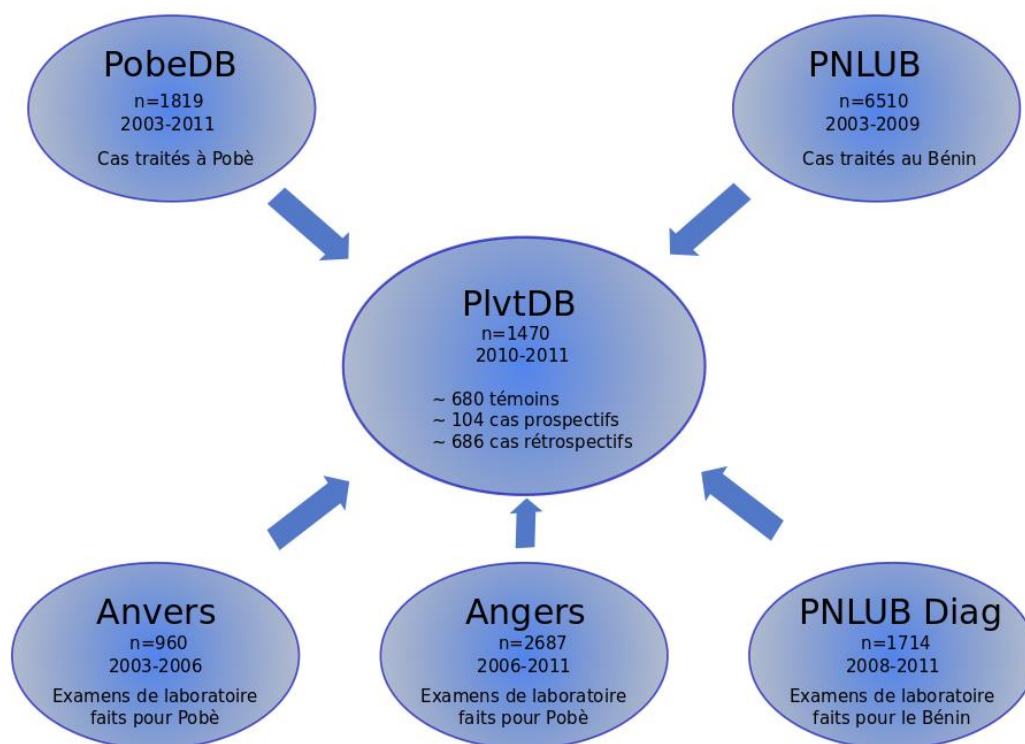
### Recrutement prospectif et rétrospectif des cas et des témoins

Le recrutement (janvier 2011 – janvier 2013) s'est déroulé parallèlement selon deux modes : prospectif et rétrospectif. Le recrutement de la cohorte primaire a pris fin en février 2012.

Le recrutement prospectif des cas est organisé sur la base suivante : tout nouveau patient présentant des signes cliniques d'UB après examen clinique par un médecin du CDTUB de Pobè, en hospitalisation ou en ambulatoire, entre janvier 2011 et janvier 2013, se voit proposer de participer à l'étude. Après information en français et/ou en fon du patient ou de son tuteur légal, un consentement éclairé est signé. Dans cette situation prospective, la participation à l'étude consiste physiquement à l'ajout d'un tube de sang (5-10 mL, tubes EDTA) à l'ensemble des prélèvements faits systématiquement au CDTUB en cas de suspicion clinique d'UB. Les informations cliniques sur l'histoire naturelle de la maladie et la confirmation par PCR du diagnostic de Buruli sont obtenues prospectivement. Le recrutement prospectif de contrôles concernait les patients examinés par un médecin du CDTUB de Pobè dans un contexte de médecine générale présentant des plaies dans le cadre de blessures accidentelles ou de pathologies non-infectieuses comme le diabète ou l'insuffisance veineuse. Environ 100 patients et 50 contrôles ont été recrutés prospectivement.

Le recrutement rétrospectif des cas et des contrôles est organisé par campagnes de prélèvement dans les dispensaires et villages endémiques autour de Pobè. Une session d'information sur l'UB et sur l'étude génétique est organisée sur la place du village. Les volontaires sont ensuite informés individuellement et prélevés après consentement. A ce stade, le statut cas-témoin est déclaratif. Le médecin ou l'infirmier pourra également constater une cicatrice d'UB ou reconnaître le patient. Environ 700 patients et 600 contrôles ont été recrutés rétrospectivement.





**Figure 29. Bases de données disponibles et assignation finale du statut cas-témoin au moment du génotypage de la cohorte primaire.**

Au moment de la constitution de la cohorte primaire pour génotypage mi-2012, 1470 prélèvements avaient été réalisés au Bénin. Le statut cas-témoin des cas rétrospectifs, initialement déclaratif, était validé par intersection de diverses bases de données de patients et d'examens complémentaires.

## Phénotype : Confirmation diagnostique des cas et des témoins

L'ensemble des individus recrutés recevaient un identifiant unique et rejoignaient une base de données des prélèvements nommée PlvtDB. Pour les cas prospectifs, le statut PCR+ des cas était complété au cours de la prise en charge. Pour les recrutés rétrospectifs, une étape supplémentaire de recouplement d'information était nécessaire pour (i) déterminer le statut PCR (+, -, non faite) du patient autodéclaré et (ii) vérifier l'absence des témoins autodéclarés des bases de déclaration des patients UB. Nous disposons de plusieurs bases de données de cas d'UB au Bénin : celle de Pobè (n=1819, 2003-2011) et celle du Plan National de Lutte contre l'Ulcère de Buruli (n=6510, 2003-2009). Nous disposons également de plusieurs bases de données de résultats de laboratoire : celle d'Anvers (n=960, 2003-2006), celle d'Angers (n=2687, 2006-2011) et celle du Plan National de Lutte contre l'Ulcère de Buruli (n=1714, 2008-2011) (Figure 29).

Cette étape de recouplement de bases de données, simple dans certains contextes, se complexifie notablement ici. Les individus recrutés ne connaissent pas leur date de naissance, déclarent un âge approximatif et n'ont pas d'adresse précise ou de numéro de sécurité sociale. Par ailleurs, l'orthographe des prénoms et noms est partiellement stable (consonnes simples ou répétées : « honnou » ou « honou » ; voyelle initiale présente ou omise : « affaton » ou « faton »...). Afin de croiser ces bases de données, j'ai détourné l'algorithme d'alignement génétique BLAST dans son implémentation Bioconductor (paquet « Biostrings » v2.31.2 (224); logiciel R v3.1.1 (172)). Cet algorithme est généralement utilisé pour aligner une courte suite de lettres composée des caractères A, C, G, T à une longue séquence de référence du génome d'une espèce donnée. Il favorise un alignement parfait, mais autorise également la présence de mutation, délétions ou insertions. Après avoir normalisé tous les noms et prénoms (pas de caractères spéciaux, pas d'accent, pas de minuscule) dans PlvtDB et dans toutes les bases cibles (utilisation d'expressions régulières, langage Perl), j'ai pu aligner un à un l'ensemble des individus prélevés avec ces bases, en tolérant au maximum 2 insertions/délétions (Figure 30). Les alignements parfaits étaient validés si le sexe,

l'ordre de grandeur de l'âge et le lieu d'habitation étaient compatibles. En l'absence d'alignement parfait, un ou plusieurs alignements imparfaits m'étaient proposés pour choix et validation manuelle.

<b>HOUNGLOBOUN ABIGAE</b>	<b>BANHONNE PAUL</b>
<b>HO _NGLOBOU _ ABIGAE</b>	<b>BANHON _E PAUL</b>

**Figure 30. Alignement des noms et prénoms par l'algorithme BLAST.**

La faible alphabétisation et l'absence de connaissance exacte de la date de naissance ont rendu nécessaire un recoupement des bases de données par des techniques non-conventionnelles. Les noms donnés ici sont des illustrations fictives.

Au final, cette procédure a permis d'assigner un statut PCR+ à 476 cas et de confirmer l'absence des bases de données d'UB de 556 témoins. Par ailleurs, ce recoupement donne accès pour chaque cas de l'étude à ses informations cliniques, en particulier à la sévérité de la maladie.

### **Gestion de la collection biologique**

Entre 5 et 10mL de sang total prélevés sur EDTA étaient transportés au CDTUB, stockés en pièce climatisée (20-25°C) puis envoyé par DHL à l'équipe de Laurent Marsollier au Centre Hospitalo-Universitaire d'Angers qui réceptionne toutes les semaines des prélèvements diagnostiques en provenance du CDTUB depuis 2006. L'extraction d'ADN est réalisée selon le protocole BACC2 (GE Healthcare) dans les jours qui suivent par le technicien du laboratoire (Camille Dantec de janvier à décembre 2011 ; Jérémie Babonneau ensuite). Au total, plus de 1600 prélèvements sanguins ont été extraits pour l'étude (entre J6 et J10 pour 50% des prélèvements). Les ADN sont redissous dans 200µL de Tris-EDTA et conservés à 4°C. Les plasmas sont conservés à -20°C. L'ensemble des ADNs est dosé une première fois au Picogreen (Quentin Vincent, Camille Dantec), mis en plaque 96 puits avec une concentration-cible de 150 ng/µL puis redosés (Quentin Vincent, Jérémie Babonneau). Tous les ADNs sont migrés par électrophorèse sur gel agarose 1% (Jérémie Babonneau).

## HumanOmni2.5-8 Product Information

Feature	Description
Number of Markers	2,379,855
Number of Samples	8
DNA Requirement	200 ng
Assay	Infinium® LCG
Instrument Support	HiScanSQ or iScan
Sample Throughput†	~1,067 samples / week
Scan Time / Sample	6.5 min (HiScan) 11.4 min (iScan)

% Variation Captured ( $r^2 > 0.8$ )	1kGP* MAF > 5%	1kGP* MAF > 1%
CEU	0.83	0.73
CHB + JPT	0.83	0.73
YRI	0.65	0.51

Data Performance	Value** / Product Specification*
Call Rate	99.65% / > 99% avg
Reproducibility	99.99% / > 99.9%
Log R Dev	0.12 / < 0.30†

Spacing	Mean / Median / 90th%*
Spacing (kb)	1.19 / 0.64 / 2.76

Marker Categories	Number of Markers*
Number of SNPs with 10kb of RefSeq genes	1,231,382
MHC / ADME	19,238 / 27,335
Non synonymous SNPs	41,900
Sex Chromosome (X / Y / PAR Loci)	55,208 / 2,561 / 418
Mitochondrial	256

† Estimate assumes one HiScan system, one AutoLoader2, one Tecan Robot, and a five-day work week.

\* Compared against June 2011 1kGP data release.

\*\* Values are derived from genotyping 470 reference samples

† Value expected for typical projects, excluding tumor samples or any samples prepared not following standard Illumina protocols.

**Tableau 7. Caractéristiques de la puce Illumina Omni2.5-8 sélectionnée pour le génotypage de la cohorte primaire.**

Document commercial Illumina.

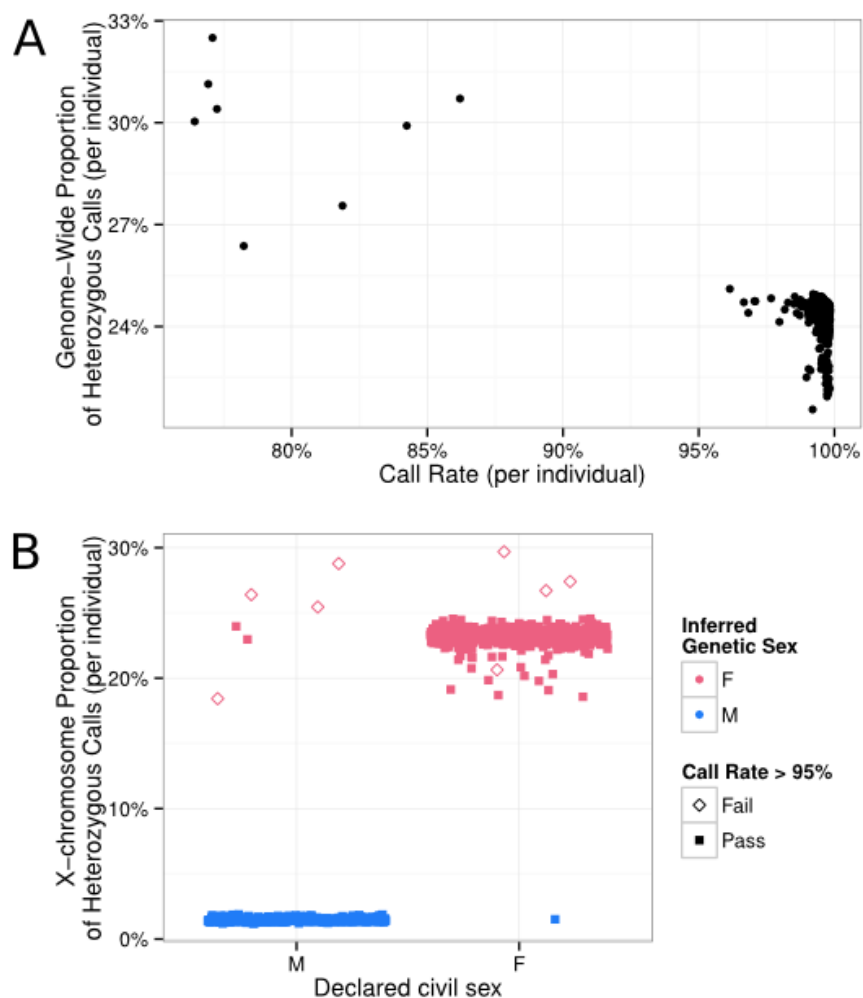
## Génotypage de la cohorte primaire

### Description de l'échantillon primaire

Un calcul de puissance a priori montre que dans un modèle additif, la puissance de détecter un SNP de fréquence 10% avec un OR de 2.1 pour un échantillon de 400 cas et 400 témoins était de 80% pour un seuil d'erreur de type I à  $10^{-5}$ . Nous avons opté pour une cohorte primaire de 408 cas PCR+ et 408 témoins recrutés entre janvier 2011 et juin 2012. Les cas VIH+ étaient exclus. Les cas, ainsi que les témoins, sont sélectionnés sur base patronymique et résidentielle pour ne pas appartenir à la même famille. Les 816 échantillons sélectionnés sont randomisés par bloc dans 9 plaques de 96 puits. Chaque bloc correspond à une demi-plaque et contient 24 cas et 24 témoins.

### Puce de génotypage

A l'été 2012, deux puces de génotypage semblent convenir à notre étude : l'Illumina Omni2.5 et l'Affymetrix Axiom Pan-AFR. Les services commerciaux déclarent respectivement que ces puces couvrent 65% et 90% des variants du génome africain (fréquence de l'allèle mineur >2%) avec un  $r^2$  de 0.8 ou plus. L'expérience passée du laboratoire, les contraintes financières et les contraintes technologiques de la plateforme partenaire ont fait préférer le génotypage sur puce Illumina Omni2.5 pour un budget d'environ 240 euros par échantillon soit 200 k€ total (Tableau 7). Le génotypage a été réalisé à la plateforme de l'Hôpital de la Pitié-Salpêtrière à Paris par Julien Guernon et Céline Méaudre sous la direction de Ioannis Theodorou. Les génotypes ont été déterminés par l'algorithme GenTrain dans l'environnement logiciel GenomeStudio d'Illumina. Les clusters (groupes) de génotypes ont été recalculés sur l'échantillon. Toutes les plaques ont été génotypées conjointement. Les résultats ont été convertis au format PLINK par un script Perl en prenant soin de supprimer les 5646 SNPs dupliqués sur la puce et d'homogénéiser l'ensemble des désignations alléliques au brin + de la séquence de référence du génome humain.



**Figure 31. Contrôle qualité par individu, n=816**

- (A) Représentation de la proportion pangénomique de génotypes hétérozygotes par individu en fonction du taux de génotypage par individu.
- (B) Représentation de la proportion de génotypes hétérozygotes sur le chromosome X par individu en fonction du sexe civil déclaré. La couleur indique le sexe inféré génétiquement et la forme indique le taux de génotypage par individu (valeur seuil de 95%).

## Contrôle qualité

### Contrôle qualité par individu

#### *Taux de génotypage – proportion d'hétérozygotes*

Le taux de génotypage dépasse 99% (des 2374209 SNPs) chez 784 individus et 95% chez 808 individus. Huit individus présentent un taux de génotypage nettement inférieur à 95%, compris entre 76 et 87%. Ces huit individus ont également un taux d'hétérozygotie (proportion de marqueurs assignés à un génotype hétérozygote) élevé (Figure 31A), ce qui suggère potentiellement un mélange d'ADN en provenance de plusieurs individus. Cette hypothèse est probable pour quatre des huit échantillons qui appartenaient à des puits adjacents (Plaque 8, Puits E1 à H1). Les individus ayant un taux de génotypage faible ont été éliminés car la fiabilité de l'ensemble des marqueurs génotypés est douteuse. Au terme de cette étape, 8 échantillons sont donc exclus.

#### *Correspondance entre le sexe déclaré et le sexe inféré génétiquement*

354 des 816 individus de la cohorte primaire étaient masculins et 462 étaient féminins. La correspondance entre le sexe civil et le sexe inféré génétiquement est un critère essentiel de fiabilité pour l'ensemble de l'échantillon (qualité du recrutement). Une absence de correspondance indique potentiellement l'inversion de plusieurs prélèvements, et remet donc en cause le phénotype enregistré. Le sexe d'un individu peut être inféré par l'analyse des marqueurs du chromosome X et Y. La méthode la plus simple consiste à mesurer le taux d'hétérozygotie du chromosome X, présent en une copie chez l'homme (haploïde, taux d'hétérozygotie attendu nul) et en deux copies chez la femme (diploïde, taux d'hétérozygotie attendu non-nul). Le taux d'hétérozygotie sur le chromosome X dans notre population sur la puce Illumina Omni2.5 était compris entre 18 et 25% chez les femmes et entre 1 et 2% chez les hommes. Le taux d'hétérozygotie était aberrant pour 7 échantillons, dont 4

déjà problématiques pour leur taux de génotypage (Figure 31B). Au terme de cette étape, 3 échantillons supplémentaires sont donc exclus.

### *Echantillons dupliqués*

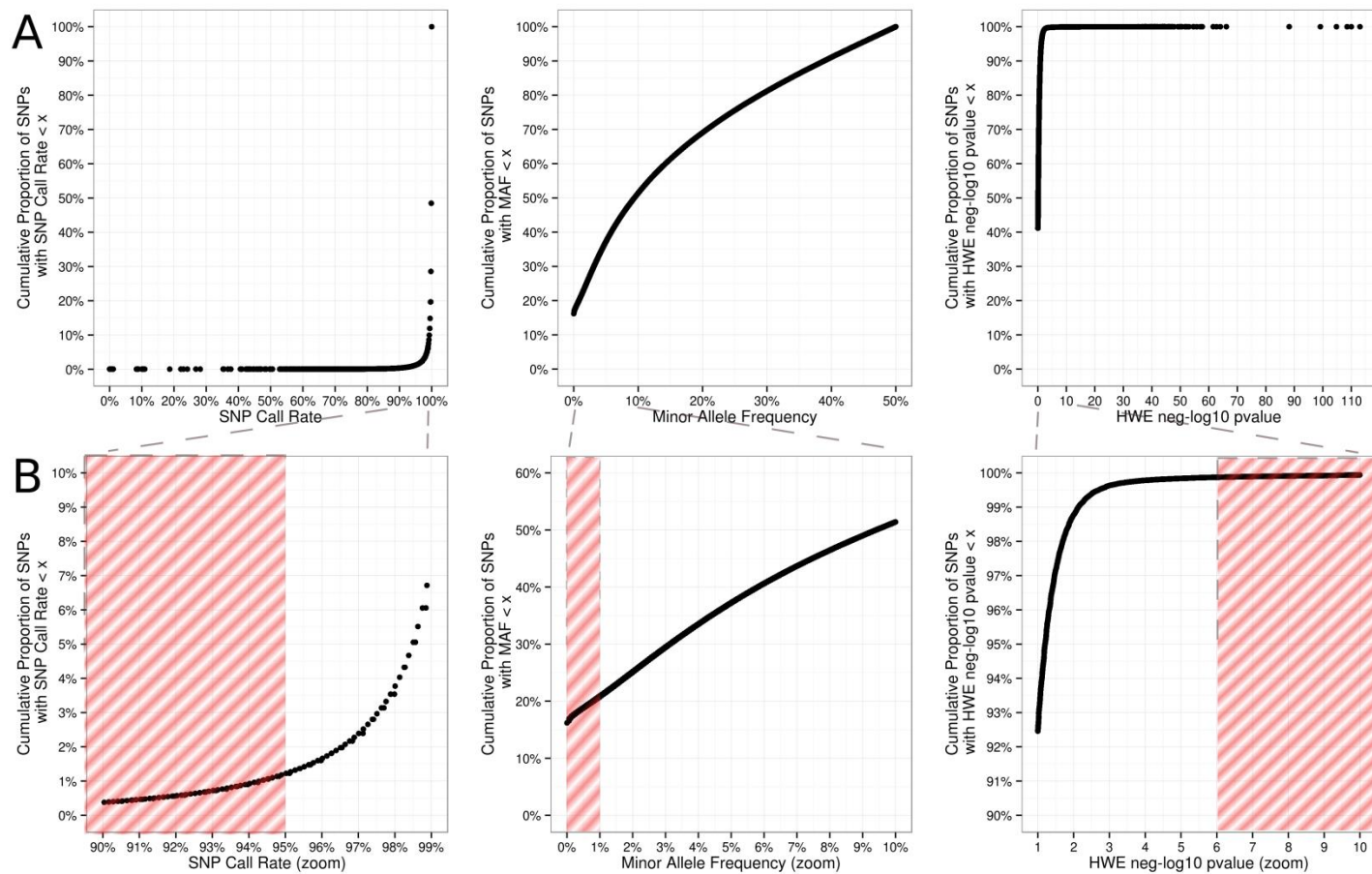
Nous avons également calculé la proportion de génotypes identiques par état (Identity by State, IBS) pour toutes les paires d'échantillons, à la recherche d'individus dupliqués. Il s'est avéré que deux paires d'échantillons avaient un IBS proche de 100%. Cette erreur dans le processus d'inclusion (deux prélèvements pour un individu) nous donne par ailleurs une estimation de la reproductibilité du génotypage fait sur deux échantillons extraits indépendamment : pour la première paire d'échantillons dupliqués, 2374162 génotypes (sur 2374209) sont concordants, et pour la deuxième paire 2372075 le sont, soit un taux de non-concordance moyen de l'ordre de  $5.10^{-4}$ , paramètre conforme aux spécifications d'Illumina. Au terme de cette étape, 2 échantillons supplémentaires sont donc exclus.

### *Composition de la cohorte primaire*

Sur les 816 échantillons génotypés dans la cohorte primaire, 8 sont éliminés pour taux de génotypage insuffisant (très inférieur à 95%), 3 sont éliminés pour incompatibilité entre le sexe civil et le sexe génétique et 2 sont éliminés pour duplication.

Après contrôle qualité individuel, l'échantillon est donc composé de 803 individus, répartis en 402 cas et 401 témoins, comprenant 348 individus masculins et 455 individus féminins. La proportion de femme parmi les cas et les témoins était respectivement de 55 et 58%. L'âge médian global était de 26 ans, celui des cas était de 14 ans et celui des témoins de 40 ans. Ceci reflète notre stratégie de recrutement non apparié sur l'âge afin de maximiser l'exposition des témoins.



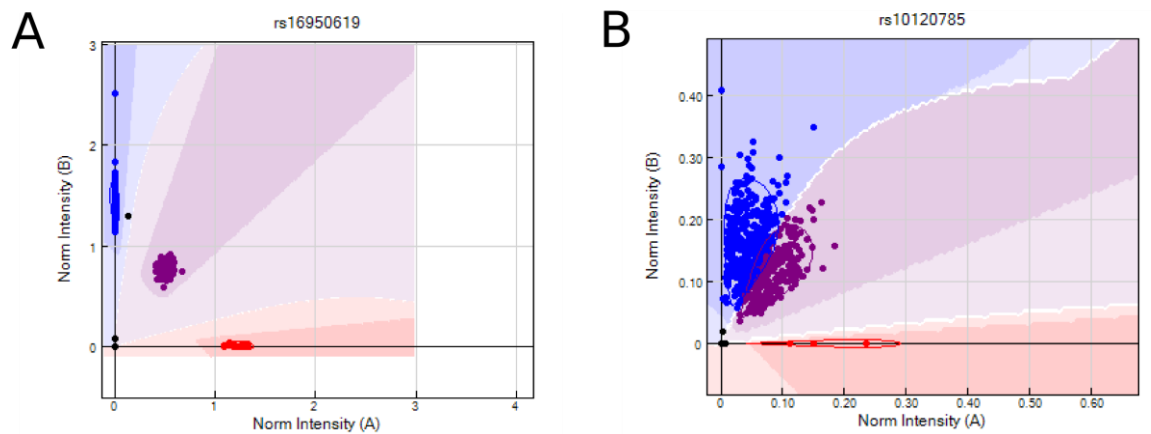


**Figure 32. Contrôle qualité par SNP, 237429 SNPs, 803 individus**

- (A) Proportion cumulée de SNPs selon leur taux de génotypage, une fréquence de l'allèle mineur et un logarithmique négatif de la pvalue du test de l'équilibre de Hardy-Weinberg.
- (B) Zooms respectifs sur les zones critiques pour le choix des seuils de filtre.

## Contrôle qualité par SNP

Le contrôle qualité par SNP est réalisé sur 803 individus. Il consiste à retirer les marqueurs dont le génotypage ne peut être considéré comme fiable.



**Figure 33. Principe de l'assignation de génotypes par puce à ADN**

- (A) Exemple d'un SNP avec assignation correcte. En abscisse, l'intensité d'hybridation normalisée pour l'allèle A et en ordonnée celle de l'allèle B. Trois groupes se distinguent clairement et permettent l'assignation d'un génotype automatisée à la quasi-totalité des individus.
- (B) Exemple d'un SNP avec assignation problématique. La difficile séparation des groupes résulte en un écart artificiel à l'équilibre de Hardy-Weinberg.

Bien que sa biochimie exacte varie d'une plateforme à l'autre, le principe du génotypage à haut-débit par puce à ADN reste constant. Dans le cas de la puce Illumina Omni2.5, le système Infinium est utilisé. L'ADN est interrogé à la position d'un SNP diallélique à l'aide de sondes ADN spécifiques de la séquence flanquante du SNP d'intérêt. L'extension de la sonde par une méthode enzymatique permet l'ajout spécifique du nucléotide complémentaire à l'allèle A ou B du SNP d'intérêt. Ce nucléotide préalablement marqué par un fluorochrome émet un signal lumineux qui est enregistré. C'est l'analyse de l'intensité des couleurs correspondant aux différents allèles du SNP qui permet la détermination du génotype de l'individu à ce SNP. Néanmoins, les intensités lumineuses émises et mesurées dépendent de nombreux paramètres (plateforme, lots de réactifs...), et

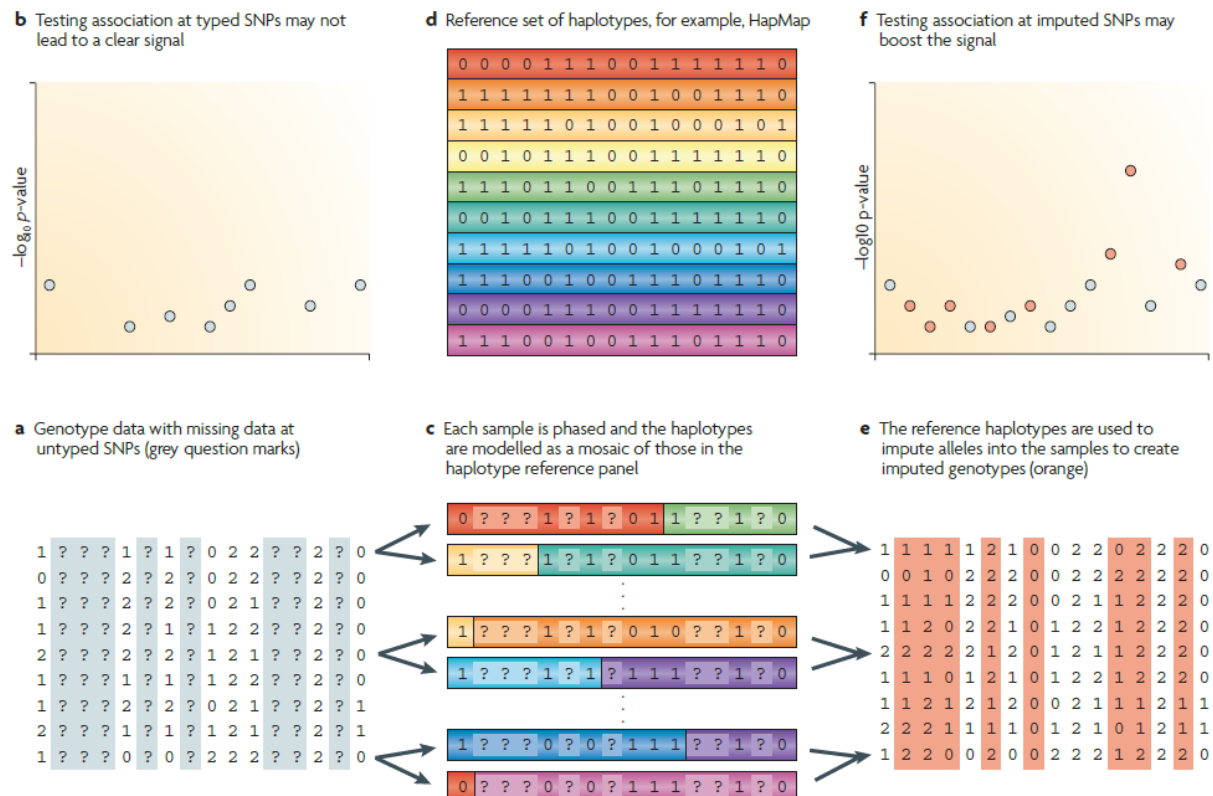
l'assignation d'un génotype se fait de manière graphique par la représentation des intensités lumineuses de chaque allèle du SNP pour l'ensemble des individus génotypés au même moment. La base des algorithmes de génotypage consiste en l'identification de 3 groupes de points correspondant aux trois génotypes possibles AA, AB et BB. Les points intermédiaires entre les groupes ne sont pas assignés à un groupe (Figure 33A).

Un échec du génotypage à un SNP peut se manifester de plusieurs façons. La répartition des intensités lumineuses peut être relativement uniforme, aboutissant à une définition peu précise des groupes, et donc un taux de génotypage faible. Par ailleurs, certaines valeurs extrêmes d'intensité peuvent tromper l'algorithme et aboutir à la fusion du groupe hétérozygote et d'un groupe homozygote. Dans ce cas, il y aura un excès d'hétérozygotes qui provoque un écart artéfactuel à l'équilibre de Hardy-Weinberg (exemple en Figure 33B). Enfin, les SNPs ayant une faible fréquence de l'allèle mineur sont caractérisés par l'absence fréquente du groupe des homozygotes pour l'allèle mineur, ce qui rend les algorithmes de génotypage moins fiables. Par ailleurs, la puissance pour détecter un effet de ces allèles dans le cadre de l'étude d'association pangénomique est faible.

Ces considérations amènent à définir les trois critères classiques de filtre de marqueurs dans le cadre d'une étude d'association pangénomique cas-témoin : le taux de génotypage (seuil généralement entre 95 et 99%), la fréquence de l'allèle mineur (seuil généralement entre 1 et 5%) et un écart majeur à l'équilibre de Hardy-Weinberg (seuil de p-value généralement entre  $10^{-4}$  et  $10^{-7}$ ). Les seuils exacts de ces filtres sont déterminés par l'observation des distributions (Figure 32).

Dans notre cas, 28022 SNPs (1.2%) avaient un taux de génotypage inférieur à 95%, 496838 SNPs (20.9%) une fréquence de l'allèle mineur inférieure à 1% et 3023 SNPs (0.13%) une p-value de test d'équilibre de Hardy-Weinberg inférieure à  $10^{-6}$  et ont été filtrés. Notamment, 384201 SNPs (16.2%) étaient monomorphes dans l'échantillon. Le panel final de SNP contenait donc 1,802,157 SNPs autosomiques.

## Box 1 | How genotype imputation works



**Figure 34. Principe de l'imputation de génotypes**

Un ensemble d'individus est génotypé pour un certains nombres de SNPs (a) dans le cadre d'une étude d'association (b). Pour imputer les génotypes inconnus, ces génotypes sont phasés (c), puis comparés à des haplotypes de référence (d) pour inférer les génotypes probables aux positions manquantes (e). Ceci peut notablement améliorer le signal d'association (f). En réalité, le tableau (e) est plus complexe puisqu'il contient pour chaque individu et chaque SNP le vecteur des probabilités des trois génotypes possibles.

*Illustration issue de (225).*

## Imputation de génotypes additionnels

L'imputation de génotypes consiste à prédire le génotype de marqueurs non directement génotypés chez un ou plusieurs d'individus (225). Un panel d'haplotypes de référence, typiquement le panel 1000 génomes, extrêmement dense en marqueurs, est utilisé pour prédire les génotypes absents d'un groupe d'individus génotypés pour un sous-groupe des marqueurs du panel de référence (Figure 34) (226). Les génotypes sont imputés sous la forme d'un vecteur de probabilité estimant la probabilité que l'individu soit AA, AB ou BB au SNP imputé sachant ses génotypes alentour et connaissant les haplotypes de référence. Plusieurs approches conceptuelles et algorithmiques à base de chaînes de Markov cachées et de méthodes de Monte Carlo par chaînes de Markov (MCMC) existent, implémentées dans différents programmes (MACH, BEAGLE, fastPHASE, IMPUTE2). IMPUTE2 semble être la méthode la plus efficace en terme de rapidité d'exécution et de qualité d'imputation (227).

Un grand gain computationnel sans perte significative de performance est obtenu en fournissant directement à IMPUTE2 les haplotypes des individus à imputer, plutôt que leurs simples génotypes (228). En l'absence de données familiales, ces haplotypes sont reconstitués par des méthodes statistiques également basées sur des chaînes de Markov cachées et des algorithmes d'estimation itératifs de type Expectation-Maximization ou MCMC (229). Plusieurs approches existent également pour cette étape (implémentées dans MACH, IMPUTE2, fastPHASE, BEAGLE, HAPI-UR, SHAPEIT2). La méthode actuellement recommandée par les auteurs d'IMPUTE2 est SHAPEIT2 (230). Cette méthode est également celle sélectionnée par le projet 1000 génomes.

Nous avons phasé nos haplotypes avec SHAPEIT2.release644 et imputé les génotypes manquants avec IMPUTE2 v2.3.0. Le phasage a été réalisé par fenêtres de 2 Mb avec les paramètres

par défaut (35 étapes MCMC, conditionnement sur 100 haplotypes, carte génétique HapMap disponible sur le site de SHAPEIT2) et la taille de population effective recommandée pour une population africaine ( $N_e=17469$ ). L'imputation a été réalisée à partir du panel de référence 1000 génomes (version du 24 août 2012) avec les paramètres par défaut et un nombre d'haplotypes de référence maximal ( $k_{hap}=2184$ , fenêtres de 5 Mb). Nous avons obtenu les génotypes de 30,071,901 SNPs autosomaux, dont environ 20 millions n'étaient effectivement polymorphes dans l'échantillon.



## Structure génétique de la cohorte primaire

Les méthodes statistiques d'estimation de l'association entre un phénotype et un génotype, qui seront présentées en détail plus loin, font l'hypothèse d'une indépendance entre chaque observation (conditionnellement au modèle). Cette hypothèse peut être violée en présence de structures génétiques qui peuvent exister à différentes échelles, comme des liens ethniques ou des liens familiaux entre individus (231,232). Il peut également exister des raisons techniques de type artéfactuelles à l'origine de ces structures génétiques, de type effet série (« batch effect ») liés par exemple à un manipulateur ou un lot de réactifs (233,234). La violation de l'hypothèse d'indépendance des observations conditionnellement au modèle (autrement dit, de l'indépendance des résidus) peut classiquement mener à une inflation de l'erreur de type I, en particulier si les structures génétiques observées sont corrélées au statut cas-témoin. Avant d'explorer les corrélations entre phénotype et génotype, il est donc nécessaire de faire le point sur la présence de structures génétiques au sein de la cohorte primaire.



## Méthodes d'identification

Plusieurs approches existent pour détecter la présence de structure dans un échantillon d'individus génotypés pour un grand nombre de marqueurs (231,235). Une première classe de méthodes dite modèle-dépendante consiste à considérer l'existence de  $K$  populations (ou groupes, ou « clusters ») dans l'échantillon, d'estimer la probabilité d'appartenance d'un individu à chaque population  $K$  via des méthodes de type MCMC et d'estimer la valeur la plus vraisemblable de  $K$  (236). Sachant  $K$  et la probabilité d'appartenance à telle ou telle population d'un individu, il est possible d'estimer les fréquences alléliques pour un marqueur donné chez les cas et témoins de chaque population, puis de formuler un test d'association stratifié, et ce de plusieurs façons (rapport de vraisemblance, test de Mantel-Haenzel, régression logistique avec ajustement sur les probabilités d'appartenance à chaque population) (231,237). Cette méthode, peut-être la plus naturelle, implémentée dans le logiciel STRUCTURE, n'est cependant pas efficace sur le plan computationnel et ne peut être appliquée à des données comportant des millions de marqueurs génétiques. Par ailleurs, l'estimation de  $K$  est problématique.

Une seconde classe de méthodes repose sur l'analyse en composante principale d'une matrice mesurant la similarité génétique entre individus. Pour résumer la ressemblance génétique de  $n$  individus sur  $m$  marqueurs (matrice  $X$  de dimension  $m \times n$  des génotypes centrés et réduits), une réduction rapide et importante de dimension est obtenue en estimant la covariance des génotypes pour chaque paire d'individu (matrice de variance-covariance  $\Sigma$ , proportionnelle à  $X^T X$ , de dimension  $n \times n$ , symétrique). Deux individus dont les génotypes sont systématiquement proches auront une valeur de covariance fortement positive ; à l'inverse deux individus dont les génotypes sont systématiquement opposés auront une valeur de covariance fortement négative. La matrice  $XX^T$  peut ensuite être reformulée comme le produit d'une matrice  $V$  de  $n$  vecteurs propres indépendants et d'une matrice diagonale  $S^2$  de  $n$  valeurs propres ou eigenvalues ( $X^T X = VS^2 V^T$ ). La matrice  $V$  contient les coordonnées de chaque individu sur les  $n$  composantes principales et la matrice  $S^2$  contient

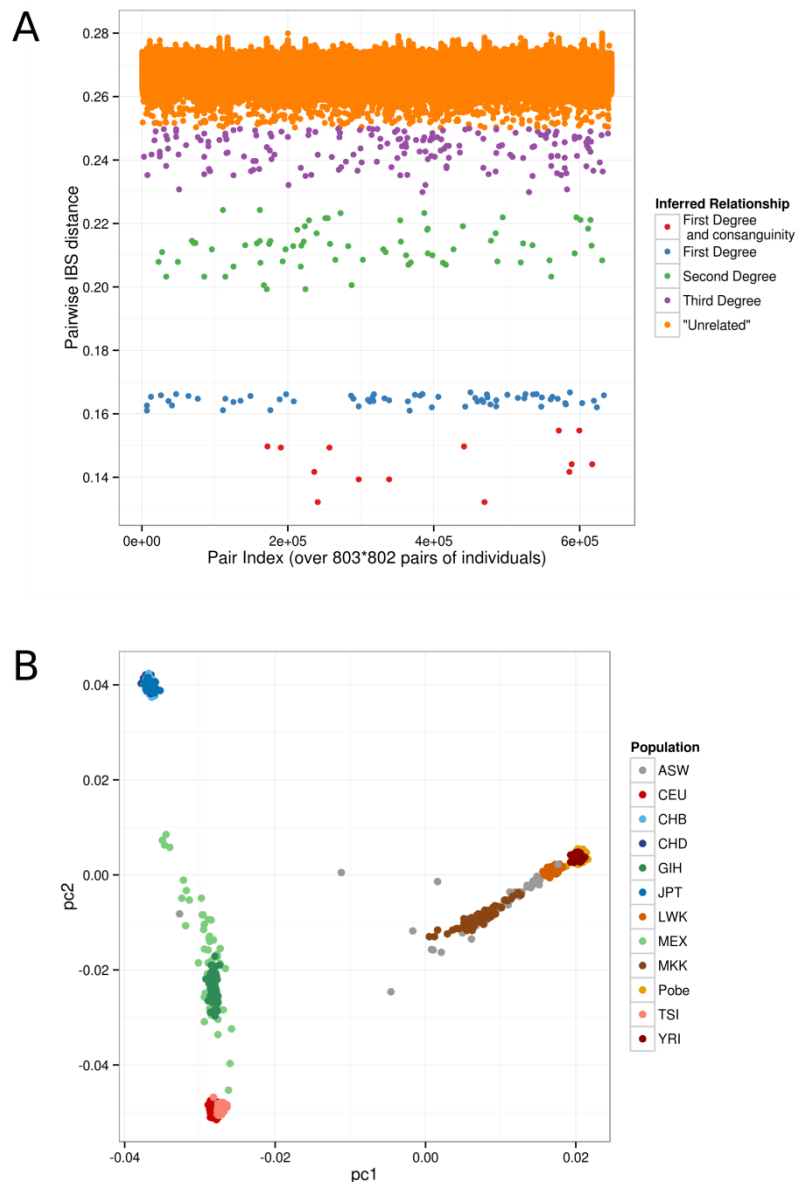
l'importance relative des différentes composantes. La matrice d'origine  $X$  peut elle-même être exprimée comme le produit  $USV^T$ , avec  $U$  la matrice de dimension  $m \times n$  des coordonnées de chaque marqueur sur les  $n$  composantes principales (également appelés poids ou loadings du marqueur sur chaque composante principale). Cette analyse en composante principale (ACP), ou décomposition en valeur singulière de la matrice  $X$ , est à la base de l'approche EIGENSTRAT (238,239).

Le principe géométrique sous-jacent à cette approche est la recherche de l'axe de projection qui conserve au maximum la dispersion des données (première composante principale), puis d'un second axe de projection orthogonal conservant au maximum la dispersion des données (seconde composante principale) et ainsi de suite. Il s'agit d'un changement de base dans un espace à  $n$  dimensions, tel que les premiers axes soient les plus informatifs. Chaque individu, initialement caractérisé par  $m$  génotypes résumés en  $n$  covariances, est maintenant caractérisé par  $n$  coordonnées sur  $n$  composantes principales, combinaisons linéaires des covariances initiales (elles-mêmes combinaisons des génotypes initiaux). Les composantes principales sont ordonnées en fonction de leur valeur propre, une mesure de la proportion de la dispersion initiale qu'elles capturent. En acceptant une perte d'information minimisée, les données peuvent donc être tronquées à  $p$  premières composantes principales, ce qui constitue la deuxième étape importante de réduction de dimension de la matrice initiale  $X$ . Le choix de  $p$  n'est pas un problème simple comme on le verra par la suite. Néanmoins, la visualisation des 10 premières composantes principales peut déjà être informative sur la structure de l'échantillon.

Autrement dit, l'information génétique relative d'un individu au sein d'un groupe peut être décrite exhaustivement par son génotype à  $m$  marqueurs ( $m \sim 10^6$ ) ou résumée par ses coordonnées sur  $p$  composantes principales ( $p \sim 10$ ). Chaque composante principale représente un axe continu de variation génétique, et sépare des groupes génétiquement homogènes : les individus appartenant à la population A auront par exemple des coordonnées négatives sur la composante principale 1 et ceux appartenant à la population B des coordonnées positives. Cette technique a été utilisée avec

succès pour reconstituer une carte géographique de l'Europe par la représentation des coordonnées de 3000 individus génotypés sur 500,000 SNPs résumés à 2 composantes principales (240).

La capacité d'une ACP à séparer des individus par groupe sur les premières composantes principales dépend fortement des niveaux de structure présents dans l'échantillon, comme nous le verrons. Ainsi, une ACP réalisée sur un échantillon composé de populations très hétérogènes (Europe et Afrique par exemple) ne séparera pas les sous-populations de chaque groupe de manière optimale (différentes populations africaines par exemple) (241). A une échelle locale, la présence de quelques individus fortement reliés, comme les membres d'une même famille, peut accaparer les premières composantes principales et dissimuler la structure génétique d'intérêt (238). Il est donc préférable d'identifier a priori les individus fortement reliés entre eux et de réaliser l'ACP sur le sous-échantillon des individus « indépendants ». Les apparentements les plus forts peuvent être identifiés par l'inspection de la matrice IBS (Identity By State) composée de la proportion d'allèles partagés par deux individus sur l'ensemble des marqueurs génotypés. Cette matrice a des liens avec la matrice de variance-covariance utilisée par EIGENSTRAT, mais n'est pas normalisée en fonction des fréquences alléliques (235).



**Figure 35. Structure génétique de la cohorte : identité par état et ACP à l'échelle mondiale**

- (A) Représentation de la matrice de distance 1-IBS (identité par état) entre chacune des 803\*802 paires d'individus. La striation apparente des valeurs de distance 1-IBS est rendue apparente par le code couleur auquel est rattaché un type de relation potentielle. Premier degré : fratrie ; parent/enfant ; second degré : grand-parent/petit-enfant, oncle/neveu ; troisième degré : cousins germains. Quelques paires d'individus aux distances IBS basses pourraient correspondre à des individus consanguins.
- (B) Représentation des deux premières composantes d'une analyse en composante principale à l'échelle mondiale incluant 803 individus de la cohorte primaire (Pobe) et 1024 individus fondateurs HapMap provenant de diverses populations mondiales. La cohorte primaire se superpose exactement à la population Yoruba. Populations d'origine africaine : Luhya du Kenya (LWK), Masai du Kenya (MKK), Yoruba d'Ibadan au Nigeria (YRI), individus afro-américains des Etats-Unis (ASW). Populations d'origine asiatique : Han de Beijing en Chine (CHB), individus d'origine chinoise de Denver aux Etats-Unis (CHD), individus de Tokyo au Japon (JPT). Populations d'origine européenne : individus caucasiens de l'Utah aux Etats-Unis (CEU), individus toscans d'Italie (TSI). Populations d'origine indienne : individus du Gujarat en Inde (GIH). Population d'origine centraméricaine : individus mexicains (MEX).



## Structure génétique de la cohorte de Pobè

### Liens familiaux cryptiques

L'examen de la matrice IBS a permis de mettre en évidence un certain nombre de liens familiaux dans l'échantillon. La Figure 35A est une représentation de cette matrice. La distance génétique entre chaque individu et les  $n-1$  autres individus de la cohorte ( $D = 1 - \text{IBS}$ ) est représentée par un point pour chaque paire d'individus distincts.  $803 \times 802$  relations sont donc successivement représentées le long de l'axe des abscisses. La valeur absolue de la distance, sur l'axe des ordonnées, dépend des fréquences alléliques des marqueurs génotypés et n'a pas d'intérêt en soi. En revanche, cette représentation rend apparente le caractère strié de la distribution observée des distances. Une inspection visuelle simple permet d'identifier 5 types de relations, avec respectivement une distance moyenne de l'ordre de 0.14, 0.165, 0.215, 0.24 et 0.27. Certains liens familiaux identifiés sur les dossiers permettent d'affirmer que la bande de distance 0.165 correspond à un lien de premier degré (coefficient de parenté  $F$  de  $1/4$ ), type parent-enfant ou frère-sœur (le coefficient de parenté entre deux individus,  $i$  et  $j$ , est égal à la probabilité pour que deux allèles tirés au hasard au même locus, l'un chez  $i$  et l'autre chez  $j$ , soient identiques par descendance (242)). Par ailleurs, la grande majorité des distances sont autour de 0.27 et représentent les individus sans liens familiaux particuliers (coefficient de parenté  $F$  nul). Un lien familial du second degré, de type grand-parent/petit-enfant ou oncle/neveu (coefficient de parenté  $1/8$ ) et un lien du troisième degré, de type cousins germains (coefficient de parenté  $F$  de  $1/16$ ) auraient respectivement une distance moyenne attendue de 0.217 et 0.244, ce qui est effectivement observé. Quelques individus avec une distance IBS faible représentent probablement des individus consanguins reliés au premier degré.

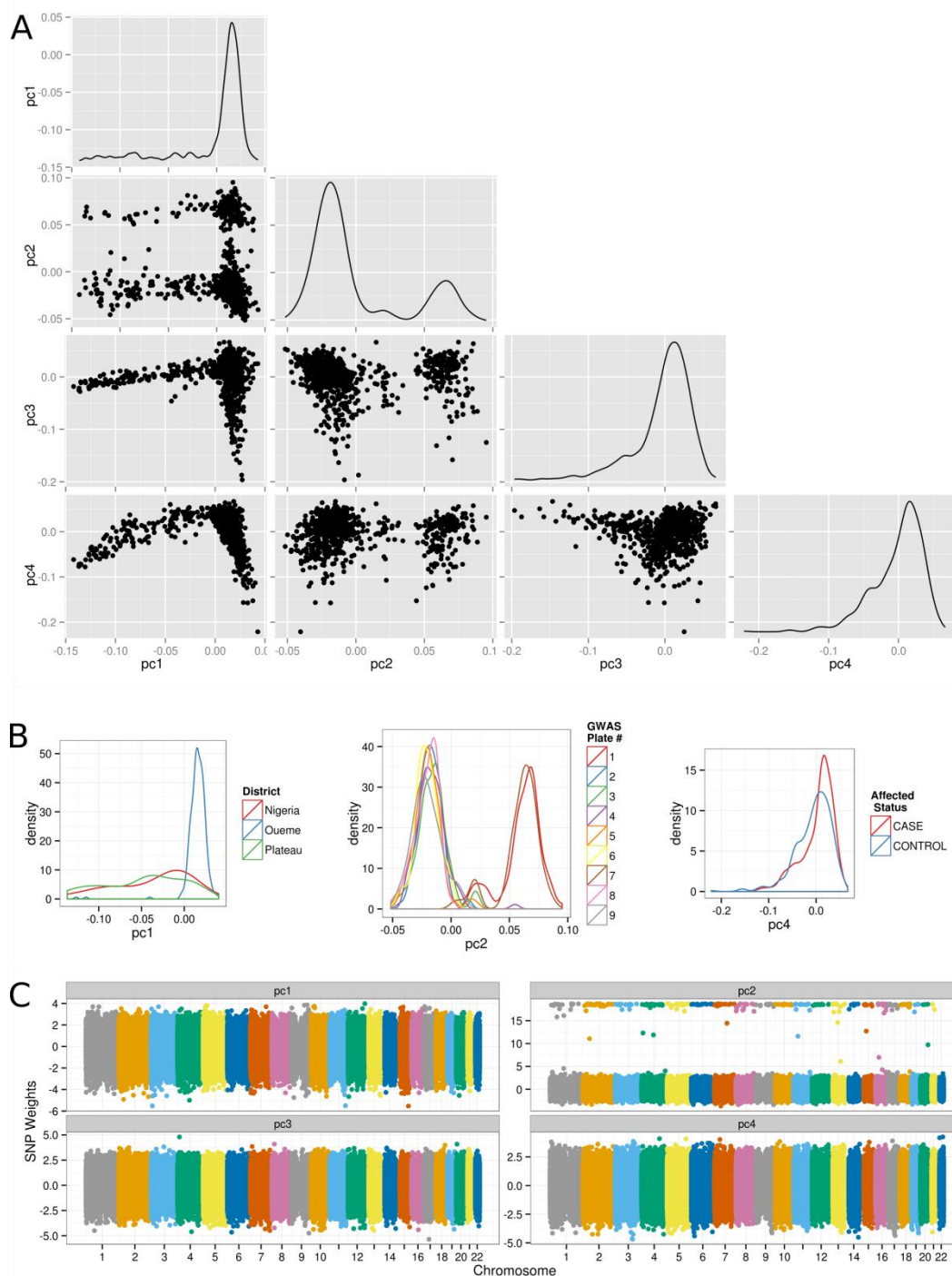
Chaque individu peut être apparenté à un ou plusieurs individus. Pour obtenir un échantillon d'individus « indépendants », on élimine itérativement  $k-1$  individus de chaque grappe de  $k$  individus. L'échantillon dit « indépendant » contient 675 individus.

## Structure de population à l'échelle mondiale

Nous avons réalisé une ACP regroupant les 675 individus indépendants de la cohorte primaire génotypés sur puce Illumina Omni2.5 et 1024 individus HapMap fondateurs génotypés sur puce Affymetrix 6.0. Les individus HapMap appartenaient à diverses populations d'origine africaine, asiatique, européenne ou américaine. Spécifiquement, les populations d'origine africaine regroupaient 90 individus Luhya du Kenya (LWK), 142 individus Masaï du Kenya (MKK), 120 individus Yoruba du Nigeria (YRI) et 53 individus afro-américains des Etats-Unis (ASW). Les populations d'origine asiatique regroupaient 90 individus Han de Chine (CHB), 90 individus d'origine chinoise des Etats-Unis (CHD), et 91 individus du Japon (JPT). Les populations d'origine européenne regroupaient 119 individus caucasiens des Etats-Unis (CEU) et 90 individus toscans d'Italie (TSI). Par ailleurs, deux populations d'origine respectivement indo-européenne et américano-européenne étaient incluses avec 89 individus du Gujarat en Inde (GIH) et 50 individus mexicains (MEX) (188,243,244).

212427 marqueurs en commun ont été identifiés. Après filtres stringents (taux de génotypage supérieur à 99%, fréquence de l'allèle mineur supérieure 5%, pvalue du test du respect de l'équilibre de Hardy-Weinberg supérieure à  $10^{-4}$ ) et élimination du déséquilibre de liaison via Plink (option `-indep-pairwise` avec des fenêtres de 50 SNPs glissantes sur 5 SNPs, seuil de  $r^2$  à 0.2), l'ACP a été réalisée sur 94697 marqueurs via l'implémentation `smartpca` du package EIGENSTRAT version 4.2 (option `outlier removal` désactivée, c'est-à-dire sans suppression des individus aux coordonnées extrêmes). Les coordonnées des 128 individus de la cohorte de Pobè non inclus dans l'ACP ont été calculées à partir de leurs génotypes et des valeurs et vecteurs propres obtenues. Ces 128 individus sont donc projetés à proximité de leurs apparentés.

La Figure 35B représente les coordonnées des 1827 individus sur les deux premières composantes principales. La cohorte de Pobè est homogène et superposée à la population HapMap Yoruba. Il n'y a donc pas de structure génétique majeure dans la cohorte de Pobè à l'échelle mondiale. Aucun individu marginal (« outlier ») n'est identifié.



**Figure 36. Structure génétique de la cohorte : ACP à l'échelle locale**

- (A)** Représentation des 4 premières composantes principales. En diagonal sont représentées les distributions des coordonnées de chaque composante chez les 803 individus sous forme de courbes de densité. Sous la diagonale sont représentées les coordonnées de chaque individu sur 2 composantes principales.
- (B)** Distribution des coordonnées des 803 individus sur la composante 1, 2 et 4 en fonction de leur facteur explicatif respectifs, le lieu d'habitation, la plaque de génotypage et le statut cas-témoin.
- (C)** Contribution de chacun des 94697 marqueurs, répartis sur 22 chromosomes, à chacune des quatre premières composantes principales.



## Structure de population à l'échelle locale

Nous avons ensuite réalisé une ACP à l'échelle locale afin de visualiser une structure génétique fine au sein de la cohorte primaire. Cette ACP incluait les 675 individus indépendants de la cohorte primaire. Comme précédemment, après filtres stringents (taux de génotypage supérieur à 99%, fréquence de l'allèle mineur supérieure à 5%, pvalue du test d'équilibre de Hardy-Weinberg supérieure à  $10^{-4}$ ) et élimination du déséquilibre de liaison via Plink (option `-indep-pairwise` fenêtres de 50 SNPs glissantes sur 5 SNPs, seuil de  $r^2$  à 0.2), l'ACP a été réalisée sur 291460 marqueurs via l'implémentation `smartpca` du package EIGENSTRAT version 4.2 (option `outlier removal` désactivée). Les coordonnées des 128 individus non inclus dans l'ACP ont été calculées à partir de leurs génotypes et des valeurs et vecteurs propres obtenues.

La Figure 36A représente les coordonnées des 803 individus sur les 4 premières composantes principales par paire de deux composantes principales. On distingue clairement à l'inspection visuelle plusieurs groupes d'individus en particulier sur les deux premières composantes principales. Une distribution bimodale de la distribution de la seconde composante principale est clairement visible sur la courbe de densité correspondante (diagonale de la Figure 36A). Il y a donc une structure génétique à l'échelle locale.

Pour tenter d'expliquer cette structure, nous avons recherché des corrélations entre chacune des dix premières composantes principales et des variables générales comme l'âge, le sexe, le lieu d'habitation, le statut cas-témoin et la localisation physique de l'ADN dans les plaques de génotypage. Pour trois composantes principales, nous avons identifié une association significative avec une variable explicative (Figure 36B). La première composante principale corrèle significativement avec la localisation microgéographique des individus, avec des valeurs négatives pour les habitants du Nigéria et du Plateau et des valeurs positives pour les habitants du Ouémé, sur les rives du fleuve. La seconde composante principale corrèle significativement avec l'appartenance à certaines plaques de génotypage, avec des valeurs positives pour les plaques 1 et 7 respectivement

et négative pour les autres. La quatrième composante principale corrèle significativement avec le statut cas-témoin. Il n'a pas été identifié de corrélat pour les autres composantes principales, avec les informations dont nous disposions.

La contribution de chaque SNP aux coordonnées des quatre premières composantes principales est représentée en Figure 36C. On observe que les SNPs contributeurs (ceux avec les poids les plus élevés en valeur absolue) sont répartis sur tous les chromosomes. Il est intéressant d'observer que pour la composante 2, artéfactuelle, les poids sont bien plus élevés en valeur absolue que pour les autres composantes principales. Il y a donc un sous-ensemble bien précis de SNPs dont les génotypes ont été biaisés sur les plaques 1 et 7.



## Analyse d'association pangénomique

Dans la suite du manuscrit, le terme « analyse d'association pangénomique » et l'acronyme GWAS (Genome-Wide Association Study) seront utilisés de façon interchangeable.

## Méthodes d'association

### Modèles génétiques : codage du génotype

L'impact d'un allèle délétère sur le risque de développer une maladie (ou symétriquement d'un allèle protecteur sur la résistance à la maladie) peut suivre différents mécanismes aboutissant à différents modèles génétiques. Un allèle récessif ne modifie le risque de maladie que s'il est présent de manière homozygote. Un allèle dominant modifie le risque de maladie dès sa présence à l'état hétérozygote et cette modification du risque est identique à l'état homozygote.

En génétique mendélienne, on trouve des explications physiopathologiques à ces modes de transmission : si la présence d'un seul allèle sauvage suffit à assurer la fonction de la protéine considérée, la maladie ne survient qu'en l'absence d'allèle sauvage et le mode de transmission est récessif. Si la présence d'un seul allèle sauvage provoque un déficit quantitatif incompatible avec la fonction de la protéine, le mode de transmission est dominant (mécanisme d'haploinsuffisance). C'est également le cas si la présence d'un allèle muté empêche l'action normale de l'allèle sauvage, par exemple par la formation de complexes protéiques non-fonctionnels (mécanisme de dominance négative) (245).

L'allèle à risque peut également augmenter le risque de maladie de façon multiplicative en fonction de sa présence à l'état hétérozygote ou homozygote. C'est l'hypothèse du modèle additif. Ce terme n'est généralement pas utilisé en génétique mendélienne, mais il existe des situations s'y rapportant : les porteurs homozygotes de la mutation du facteur V de Leiden ont un risque thrombotique multiplié par 80, tandis que les hétérozygotes ont un risque multiplié par 7 (la transmission est toutefois classiquement qualifiée de dominante) (246). Le surrisque de maladie peut également n'exister qu'en cas d'homozygotie, que ce soit pour l'allèle sauvage ou l'allèle à risque

(situation dite « d'avantage aux hétérozygotes »). Ce type de mécanisme est documenté en génétique animale et végétale, et également observé chez l'homme, tout particulièrement dans le domaine des maladies infectieuses où il pourrait être associé à de meilleures capacités de reconnaissance des pathogènes (247–253).

Modèle Génétique	Risque Relatif Génotypique		
	Génotype AA	Génotype AB	Génotype BB
Récessif	1	1	$\gamma_{BB} \neq 1$
Dominant	1	$\gamma_{AB} = \gamma_{BB} \neq 1$	$\gamma_{BB} = \gamma_{AB} \neq 1$
Additif / Allélique	1	$\gamma_{AB} = (\gamma_{BB})^{0.5} \neq 1$	$\gamma_{BB} = (\gamma_{AB})^2 \neq 1$
Avantage aux hétérozygotes	$\gamma_{AA} = \gamma_{BB} \neq 1$	1	$\gamma_{BB} = \gamma_{AA} \neq 1$
Génotypique	1	$\gamma_{AB} \neq 1$	$\gamma_{BB} \neq 1$

**Tableau 8. Risques relatifs génotypiques de chaque modèle génétique utilisé en génétique complexe.**

Un risque relatif génotypique  $> 1$  implique que le génotype est à risque et un risque relatif génotypique  $< 1$  implique que le génotype est protecteur.

La génétique complexe a repris les modèles hérités de la génétique mendélienne et les a formalisés en fonction des risques relatifs génotypiques (Tableau 8). Néanmoins, le modèle le plus populaire pour les études d'association pangénomique est nettement le modèle additif. Il est difficile de trouver l'origine de cet état de fait. Le choix du modèle additif n'est pas justifié dans un article standard (254). Le registre des associations phénotypes-génotypes (« GWAS catalog ») maintenu par le National Human Genome Research Institute (NHGRI) ne retient d'ailleurs que les p-values et odds-ratios du modèle allélique, qui sont asymptotiquement égaux aux p-values et odds-ratios du modèle additif en cas d'équilibre de Hardy-Weinberg (223,255).

Les arguments apportés pour ne tester (officiellement) qu'un seul modèle génétique sont les suivants (256) : (i) le modèle additif est asymptotiquement équivalent en cas d'équilibre de Hardy-

Weinberg au modèle allélique qui compare directement la fréquence de l'allèle considéré chez les cas et les témoins, ce qui simplifie la description de l'allèle comme « facteur de risque » au sens de l'épidémiologie classique, et ce qui simplifie également la conduite de méta-analyses; (ii) le modèle additif serait relativement neutre et pourrait capturer d'autres modèles vrais ; (iii) si le variant étudié n'est pas le variant causal mais un marqueur en déséquilibre de liaison avec lui, la puissance de détection d'un effet réellement non-additif du variant causal serait minimale au variant étudié dès lors que le déséquilibre de liaison n'est pas complet (256) ; (iv) les questions des tests multiples et du choix des marqueurs à géotyper dans la cohorte de réplication sont bien assez difficiles à gérer avec un seul modèle génétique.

La génétique mendélienne montre qu'il n'y a pas de justification biologique à une préséance d'un modèle additif. Des études de simulation et des dérivations analytiques rappellent d'ailleurs que le modèle le plus puissant pour la détection de l'effet d'un variant est le modèle « vrai », que nous ne connaissons pas (257,258). L'idée d'une « neutralité » du modèle additif vient du fait qu'il modélise un risque à l'état hétérozygote (comme le modèle dominant et le modèle avantage aux hétérozygotes) et à l'état homozygote (comme le modèle dominant et le modèle récessif). Il impose néanmoins une relation multiplicative entre l'état hétérozygote et homozygote. Sa neutralité est donc partielle ; un variant à effet récessif est particulièrement difficile à capturer par ce biais (257,258).

Il existe d'ailleurs un modèle plus neutre que le modèle additif, le modèle génotypique, qui n'impose aucune relation entre l'état hétérozygote et homozygote. Ce modèle requiert néanmoins l'estimation de deux paramètres de risque, alors que tous les autres modèles aboutissent à l'estimation d'un seul paramètre de risque. Encore une fois, ce modèle est le plus puissant s'il correspond au modèle vrai. C'est le modèle le plus universel, puisqu'en cas de modèle vrai récessif, l'estimation du risque pour les hétérozygotes tendra vers 1 et qu'en cas de modèle vrai dominant, les estimations des risques pour les hétérozygotes et les homozygotes tendront à être égales. En cas de

modèle vrai additif ou avantage aux hétérozygotes, le modèle génotypique tendra également vers les modèles vrais. Mais les tests statistiques dérivés auront un degré de liberté supplémentaire, ce qui aboutira à une perte de puissance dans toutes les situations où le modèle vrai est additif, récessif, dominant ou avantage aux hétérozygotes. Le coût statistique moyen de ce paramètre supplémentaire est à l'origine de la rare adoption de ce modèle génétique. La célèbre étude du Wellcome Trust Case Control Consortium a cependant choisi de tester systématiquement deux modèles, le modèle additif et le modèle génotypique. Les modèles dominants et récessifs ne sont considérés que si le modèle génotypique améliore significativement l'adéquation aux données (model fit) (259).

Il y a un certain pragmatisme à ne traiter que le modèle additif. Pour beaucoup de marqueurs, la fréquence des homozygotes pour l'allèle mineur est faible, ce qui rend en pratique l'estimation d'un OR pour un effet dominant ou un avantage aux hétérozygotes proche de l'OR estimé par le modèle additif. Le seul modèle réellement négligé par cette approche est le modèle récessif. Certains considèrent que la puissance pour détecter un effet récessif dans un GWAS est trop faible pour que l'on s'y attarde. A odds-ratio et fréquence allélique constants, la puissance de détection d'un effet récessif est bien plus faible que celle d'un effet dominant, en raison de la rareté globale des cas et témoins exposés dans le cas récessif (homozygotes). Enfin, on pourrait argumenter que la capacité technologique ou même conceptuelle de validation mécanistique d'un signal de GWAS est bien trop faible pour que le mode réel d'action de l'allèle à risque ait une importance à l'heure actuelle. Qui plus est, tester un seul modèle réduit la dimension du problème de la correction pour tests multiples.

Il reste une constatation simple : nous sommes des êtres diploïdes et la biologie mendélienne regorge d'exemples de mode d'action de toute nature, en particulier récessifs et dominants. La question des tests multiples ne nous paraît pas centrale dans une perspective exploratoire de génération d'hypothèses. Nous avons donc choisi d'explorer l'ensemble des modèles génétiques.

## Modélisation du lien phénotype-génotype

### *Modèle linéaire généralisé de type régression logistique*

Le design cas-témoin de notre étude oriente naturellement la méthode d'analyse vers la régression logistique, c'est-à-dire un cas particulier de modèle linéaire généralisé dans lequel la probabilité de développer la maladie est modélisée par une combinaison linéaire de facteurs explicatifs que sont le génotype (codé selon différents modèles génétiques, voir supra) et des variables d'ajustement. Plus précisément, le logarithme de la cote de la probabilité de développer la maladie est modélisé (fonction de lien logit). Ce modèle linéaire est adapté à l'estimation de probabilités dont les valeurs sont par essence contraintes à l'intervalle  $[0,1]$ . Le choix de modéliser la cote rend par ailleurs aisée l'estimation des odds-ratios (rapport de cote) quantifiant le risque conféré par tel ou tel génotype comme l'exponentielle du coefficient de régression linéaire. La formulation d'une fonction de vraisemblance des données en fonction de ce coefficient permet son estimation au maximum de vraisemblance. Le caractère significativement différent de 0 de ce coefficient est évalué par un test reposant sur la statistique du rapport de vraisemblance. L'ensemble de ces procédures a été réalisé dans le logiciel R par la fonction `glm` (172). La prise en compte d'une structure génétique de l'échantillon de type stratification de population, ou effet batch, peut être réalisée par l'introduction des composantes principales (issues de la décomposition de la matrice de variance-covariance des génotypes entre individus) comme variables d'ajustement dans le modèle de régression logistique. Nous avons réalisé l'analyse sans et avec ajustement sur les premières composantes principales. On verra que le nombre de composantes principales à inclure est objet à débat scientifique.



### *Modèle linéaire mixte*

La prise en compte de liens familiaux cryptiques n'est pas assurée par l'ajustement sur les composantes principales, en particulier quand celle-ci a été réalisée sur le sous-échantillon des individus « indépendants » pour permettre une meilleure détection des populations. Comme la stratification de population, les liens familiaux cryptiques peuvent provoquer une inflation de l'erreur de type I (232,233). Leur prise en compte est possible dans le cadre de modèles linéaires mixtes incorporant une matrice de similarité génétique entre individus, comme une matrice IBS ou une matrice de variance-covariance des génotypes normalisés. Les modèles mixtes sont très populaires dans le domaine de l'analyse de données répétées (plusieurs mesures chez un individu à différents temps, par exemple), où ils permettent la prise en compte naturelle d'un effet « sujet » en attribuant à chaque individu une ordonnée à l'origine (intercept) propre, issue d'un tirage d'une variable aléatoire dont on modélise la variance. Dans le cadre qui nous intéresse, il n'y a pas de données répétées à strictement parler, puisque chaque individu n'est représenté que par un couple {phénotype, génotype}. Néanmoins, les groupes d'individus appartenant à une même famille ne représentent pas des données indépendantes et s'apparentent à ces données répétées. Le modèle mixte modélise l'effet du marqueur sur le phénotype comme un effet fixe, à l'instar de la régression linéaire classique, mais modélise les erreurs résiduelles comme issues du tirage d'un vecteur aléatoire dont la structure de covariance est contrainte par la spécification d'une matrice de similarité génétique. Autrement dit, le modèle attribue à chaque individu une ordonnée à l'origine qui lui est propre en contraignant les individus proches à avoir des ordonnées à l'origine proches. Notons que l'estimation précise du lien entre individus nécessite des données pangénomiques et ne peut concerner que l'échantillon de la cohorte primaire. Une étude récente montre que l'estimation de la matrice de similarité sur un faible nombre de SNPs (environ 2000) affecte significativement l'efficacité de la prise en compte des liens familiaux cryptiques en termes d'erreur de type I (260).

Les méthodes d'estimation sont plus complexes pour les modèles mixtes que pour les modèles linéaires classiques, et demandent en règle générale de grandes ressources computationnelles. Pour rendre le modèle mixte accessible aux approches pangénomiques où des millions de modèles sont évalués, plusieurs implémentations ont été proposées, comme GRAMMAR, EMMA, EMMAX, FAST-LMM (260). L'accélération proposée par GEMMA ne repose pas sur des approximations mais sur une réécriture matricielle efficace du problème d'optimisation (261). Une récente étude comparative montre qu'en pratique ces différentes implémentations donnent des résultats très similaires (260). Il est intéressant par ailleurs de noter que l'algorithmique des modèles mixtes dans ce cadre passe par l'analyse en composantes principales de la matrice de similarité génétique, ce qui sous-entend certains ponts entre les différentes approches. Enfin, notons qu'aucune de ces méthodes n'implémente de modèle mixte généralisé de type modèle logistique et qu'un phénotype binaire est modélisé comme le serait un phénotype continu. GEMMA ne comporte par défaut que l'analyse en modèle additif. Nous avons développé un script perl permettant la transformation des données pour les modèles dominant, récessif et avantage aux hétérozygotes. Le modèle génotypique n'a pas été exploré dans cette approche.

### *Prise en compte de l'âge de survenue de la maladie par analyse de survie*

L'âge de survenue apporte une information importante sur l'architecture de la contribution génétique dans la genèse de la maladie (262). Le développement d'une maladie dite mendélienne est en effet typiquement déterminé par une lésion génétique majeure dans un gène central, tandis que le développement d'une maladie complexe est simplement rendu plus probable par la présence conjointe de plusieurs allèles à risque dans un environnement donné. Or les maladies mendéliennes sont à développement généralement plus précoce que les maladies complexes. En suivant ce même raisonnement, on peut supposer qu'au sein de patients atteints par un phénotype complexe, le développement précoce du phénotype indique une contribution génétique importante. A l'inverse,

on peut supposer que le développement tardif de la maladie est le fruit d'un processus physiopathologique multifactoriel et multi-étape particulièrement complexe. Il semble donc pertinent de prendre en compte l'âge de survenue de la maladie dans l'analyse. Cet âge de survenue peut être modélisé dans le cadre de l'analyse de survie.

L'analyse de survie permet l'étude de la survenue de tout événement binaire absorbant (typiquement le décès, d'où son nom) et s'applique donc à la survenue d'une maladie. La modélisation classique des données de survie se fait au sein du modèle des risques proportionnels de Cox. Dans notre cas, l'exposition d'intérêt est le génotype à un marqueur donné, l'événement d'intérêt le développement de la maladie, le délai de survenue est l'âge au développement de la maladie et le délai censuré de survie est l'âge des témoins à l'inclusion. Cet âge représente effectivement une durée d'exposition puisque les villages autour de Pobè sont hyperendémiques pour l'UB et que les flux migratoires sont quasi-inexistants. L'estimation du risque absolu de développer la maladie est impossible dans notre étude, puisque les cas sont surreprésentés par rapport à la population générale. En revanche, les témoins n'ont volontairement pas été appariés sur l'âge ou le sexe, et représentent en première approximation un échantillon aléatoire de la population générale. L'analyse de survie pour notre échantillon a été réalisée sous R, paquet « survival » (263).

Enfin, la question des liens cryptiques entre individus et de leur non-indépendance se pose également pour l'analyse de survie et peut également être résolue par l'utilisation d'une matrice de similarité génétique au sein d'un modèle mixte. En pratique, l'analyse est extrêmement lente. Nous avons donc réalisé cette analyse sur l'échantillon des 12091 meilleurs SNPs issus des autres modèles statistiques, qui correspondait à l'union de la liste de 2500 meilleurs signaux de chaque analyse et de la liste de tous les signaux inférieurs en p-value à  $10^{-3}$ . Cette analyse de survie a été réalisée sous R, paquet « coxme » (264).

### *Cas des SNPs imputés*

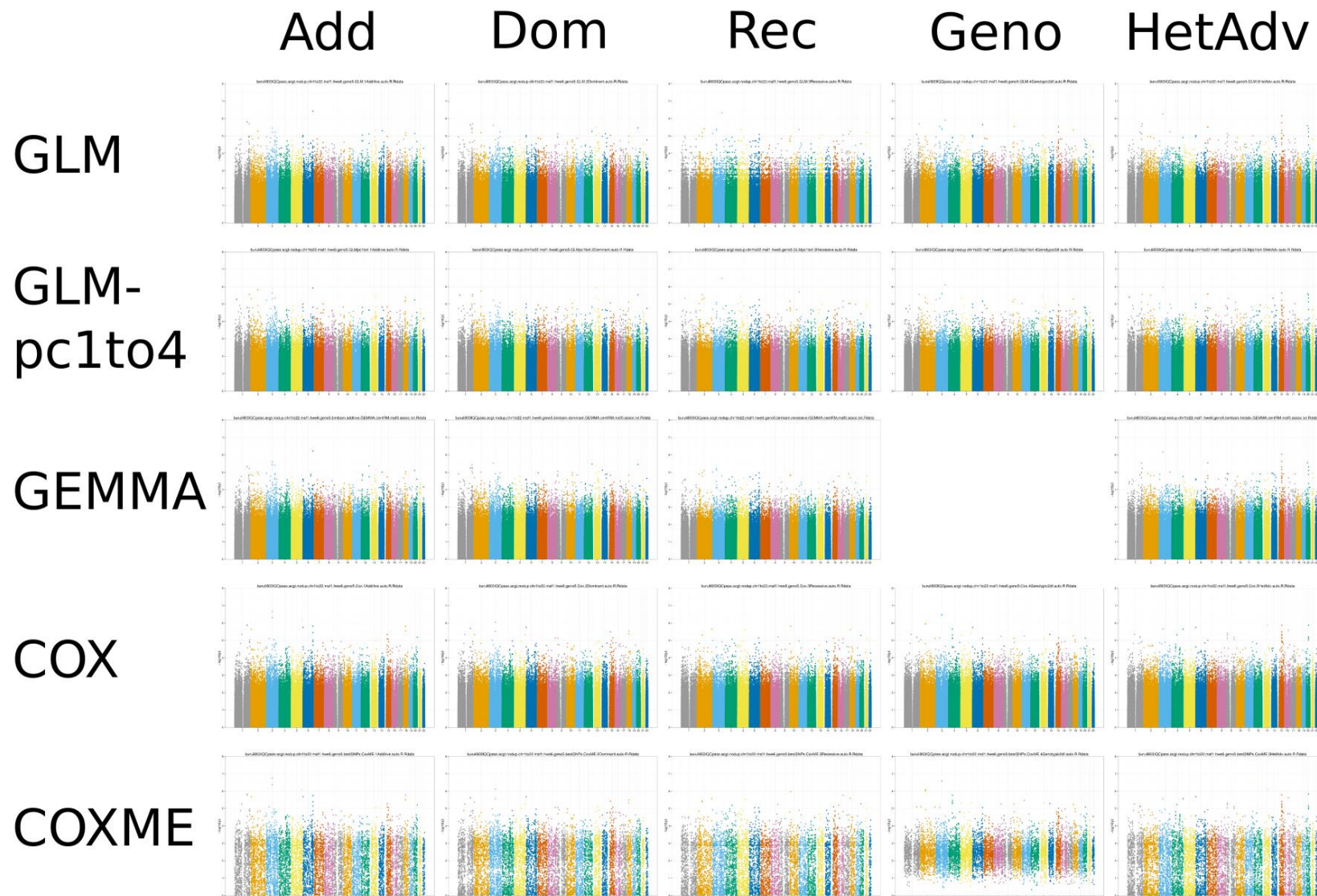
Dans le cas des SNPs imputés, la contribution de chaque individu n'est pas un couple {Phénotype ; Génotype} mais un quadruplet { Phénotype ; Probabilité du Génotype AA,  $P_{AA}$ ; Probabilité du Génotype AB,  $P_{AB}$  ; Probabilité du Génotype BB,  $P_{BB}$  }. La prise en compte de cette incertitude dans le modèle de régression logistique peut suivre plusieurs approches, toutes implémentées dans le logiciel SNPtestv2.4 (225,265).

L'approche la plus simple est d'assigner un génotype à chaque individu lorsque la probabilité d'un des trois génotypes dépasse un seuil donné, par exemple 90%. On rejoint alors le cas général d'un SNP génotypé. Cette approche peut néanmoins aboutir à un fort taux de génotypes non assignés.

Une deuxième approche consiste à calculer l'espérance de l'exposition de chaque individu à partir du vecteur de probabilité des génotypes. Dans un modèle additif, c'est le nombre attendu d'allèle B auquel est exposé l'individu, qui n'est plus un entier {0,1,2} mais une variable continue bornée appartenant à l'intervalle [0,2] et valant  $P_{AB} + 2 * P_{BB}$ . Ce nombre est également appelé dosage en allèle B. On réalise ensuite la régression logistique du phénotype sur ce dosage. Dans le cas d'un modèle dominant par exemple, ce dosage est  $P_{AB} + P_{BB}$ . Cette approche est séduisante mais consiste encore en une perte d'information (puisque par exemple plusieurs combinaisons de vecteurs de probabilités pourraient aboutir à un même dosage). L'interprétation des odds-ratios estimés est délicate dans le sens où un individu porte en réalité un génotype donné discret, et non un dosage de génotype.

La troisième approche permet d'éviter cet écueil en incorporant l'incertitude autour du génotype dans le modèle de régression. La contribution de chaque individu à la vraisemblance globale devient la somme de sa contribution par génotype pondérée par la probabilité de ce génotype. L'estimation des paramètres du modèle est plus complexe que dans les autres situations et peut suivre différentes approches implémentées dans SNPtest. Nous avons utilisé l'option Maximum Likelihood et SNPtestv2.4. Les SNPs imputés avec une information supérieure à 0.6 ont été analysés.





**Figure 37. Vue aérienne de Manhattan : analyses pangénomiques d'association de l'UB**

Les 24 modèles testés sont représentés (voir Annexe pour agrandissements). Modèles génétiques : additif (Add), Dominant (Dom), Récessif (Rec), Génotypique (Geno) et Avantage aux Hétérozygotes (HetAdv). Modèles statistiques : modèle linéaire généralisé de type régression logistique (GLM), avec ajustement sur les quatre premières composantes principales (GLM-pc1to4), modèle linéaire mixte (GEMMA), modèle de Cox (Cox), modèle de Cox mixte (CoxME). Les analyses ont été réalisées sur 1,802,157 SNPs sauf pour CoxME (sélection des 12091 meilleurs SNPs des autres analyses). Il n'y a pas d'implémentation du modèle génotypique dans GEMMA.

## Analyse de la cohorte primaire

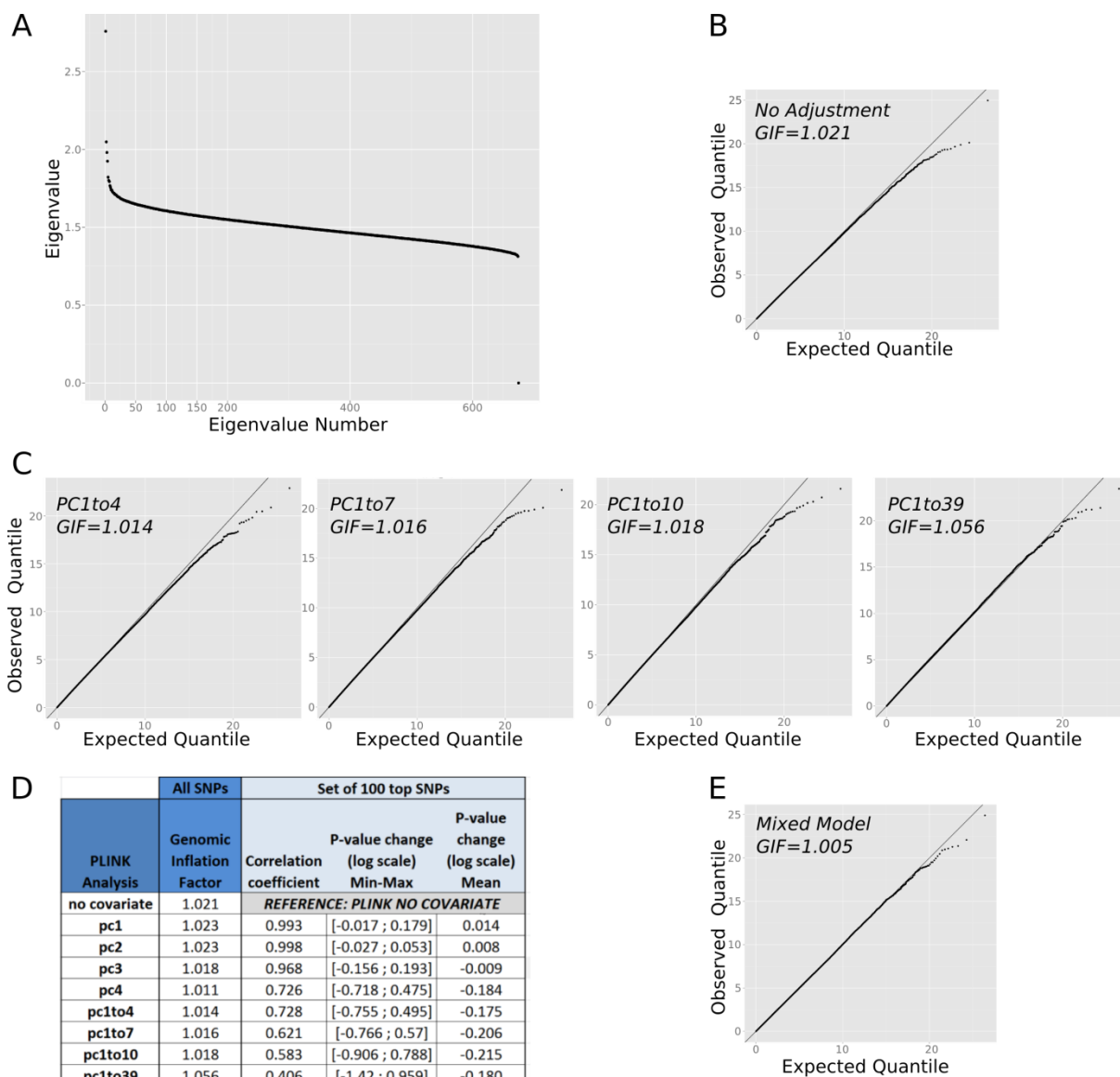
### Résultats préliminaires

Dans le cas des SNPs génotypés, nous avons donc analysé nos données selon cinq modèles génétiques (additif, dominant, récessif, avantage aux hétérozygotes, génotypique) et cinq modèles statistiques (modèle linéaire généralisé de type régression logistique sans et avec ajustement sur les composantes principales, modèle linéaire mixte incorporant une matrice de similarité génétique, modèle des risques proportionnels de Cox, et modèle des risques proportionnels mixte). Le modèle génotypique n'a pas pu être réalisé simplement sous GEMMA en modèle mixte. Le modèle de Cox de type modèle mixte n'a été réalisé que sur 12091 SNPs représentant l'union des 2500 meilleurs signaux des autres modèles. Au total, 19 à 24 analyses ont donc été réalisées sur chaque SNP.

La cohorte primaire comprenait 402 cas et 401 témoins. Après contrôle qualité, les analyses d'association étaient réalisées sur 1,802,157 SNPs autosomaux. Une représentation des résultats des 24 modèles testés est donnée en Figure 37. Il s'agit de « Manhattan plots », des graphes représentant la p-value (plus précisément, le logarithme négatif de la p-value, par exemple 5 pour  $10^{-5}$ ) de chaque SNP le long des 22 chromosomes autosomaux. Cette figure illustre la masse de données générées par les différentes analyses. On peut constater sur les Manhattan Plot en Annexe que ces différentes approches donnent des résultats corrélés mais distincts. La proportion d'hommes et de femmes était identique parmi les cas et les témoins ; l'ajustement ne modifiait pas les résultats.

Avant de se pencher sur les signaux d'association en eux-mêmes, il est important de s'assurer qu'aucune des approches utilisées n'aboutit à une inflation démesurée de l'erreur de type I. Nous avons vu dans la partie précédente que la cohorte primaire possédait une structure génétique à l'échelle locale, liée à la microgéographie locale, à un effet artéfactuel de type effet batch et au statut cas-témoin. Par ailleurs, l'analyse de la matrice IBS a dévoilé une striation caractéristique de liens familiaux cryptiques dans l'échantillon.





**Figure 38. Choix du nombre de composantes principales à inclure pour ajustement**

- (A) Distribution des valeurs propres de l'ACP. Visuellement, une sélection naturelle du nombre de composantes à inclure pourrait être 1, 4 ou 7. EIGENSTRAT propose 10 par défaut. La statistique de Tracy-Widom appliquée à nos données propose 39 au seuil de 5%.
- (B) Diagramme Quantile-Quantile de l'analyse de 1802157 SNPs avec un modèle linéaire généralisé de type régression logistique et un modèle génétique de type additif.  
*GIF : Facteur d'Inflation Génomique.*
- (C) Diagrammes Quantile-Quantile pour diverses stratégies d'ajustement.
- (D) Corrélation entre l'analyse sans ajustement et diverses analyses avec ajustement pour les 100 meilleurs SNPs de l'analyse sans ajustement. L'étendue et la moyenne des écarts entre les p-values originales et après ajustement sont données (échelle logarithmique).
- (E) Diagramme Quantile-Quantile pour le modèle linéaire mixte (GEMMA).

## Prise en compte de la structure génétique de la cohorte primaire : choix du nombre de composantes principales d'ajustement

Pour évaluer l'impact des structures génétiques identifiées sur l'erreur de type I, nous pouvons tracer les diagrammes quantile-quantile ou Q-Q plots pour chaque modèle testé. Ces graphes représentent pour chaque SNP la valeur de la statistique observée dans l'échantillon en fonction de la valeur attendue de la statistique sous l'hypothèse nulle. Par exemple, après classement par ordre croissant de la statistique de test sur l'ensemble des SNPs testés, le 180000<sup>ème</sup> SNP sur 1800000 SNPs (10<sup>ème</sup> percentile) a une valeur attendue de Chi2 à 1 degré de liberté de 0.016. Cette valeur et la valeur observée sont représentées. Une inflation de l'erreur de type I est probable si le ratio de la valeur observée sur la valeur attendue est supérieur à 1. Autrement dit, pour une statistique contrôlant l'erreur de type I, le diagramme quantile-quantile est la droite  $y=x$ . Une façon fréquente de résumer un diagramme quantile-quantile est de donner le ratio de la statistique observée sur la statistique attendue à la médiane de la distribution (valeur attendue 0.45 pour un chi2 à 1 degré de liberté). Dans le cadre des analyses d'association pangénomique, ce ratio est nommé facteur d'inflation génomique (*Genomic Inflation Factor*, GIF). Il est à noter que l'ensemble de cette procédure est valable sous l'hypothèse que la majorité des SNPs n'est pas associée à la pathologie d'intérêt et que les statistiques obtenues sont bien distribuées selon l'hypothèse nulle.

La question du nombre de composantes principales sur lequel ajuster le modèle de régression est une question épineuse et sans réponse universelle. Il s'agit de sélectionner les composantes principales représentant des « vraies » structures génétiques pouvant jouer le rôle de facteur de confusion dans le lien entre le variant étudié et le phénotype d'intérêt. L'équipe ayant introduit l'ACP dans le domaine de l'analyse d'association pangénomique propose implicitement trois approches (233,238). La première, implémentée par défaut dans le package EIGENSTRAT, ne consiste pas en un ajustement, mais au retrait des individus marginaux (« outliers ») de l'échantillon d'analyse. C'est l'approche qui a été utilisée par le Wellcome Trust dans son papier-référence (259).

Une seconde approche est de tester formellement le nombre de valeurs propres statistiquement significatives en s'appuyant sur la distribution théorique de la valeur propre la plus élevée développée par Tracy et Widom (238). Pour des raisons qui restent à déterminer, cette approche n'a eu que peu de succès en pratique et n'est d'ailleurs pas reprise dans le papier de revue publié par les mêmes auteurs (233). La troisième possibilité est de garder la valeur définie par défaut de 10 composantes principales, dans l'hypothèse que cela devrait couvrir la plupart des situations pratiques de stratification de population. Une possibilité moins arbitraire est l'inspection visuelle de la distribution des valeurs propres associées aux composantes principales. Ces valeurs-propres quantifient l'importance relative de chaque composante et fournissent parfois un seuil naturel. D'autres ont choisi de n'ajuster que sur les composantes principales qu'ils ont pu relier effectivement à des populations identifiées culturellement (266). On pourrait également choisir de n'ajuster que sur les composantes principales corrélant avec le statut cas-témoin, puisqu'il s'agit d'une condition nécessaire pour la qualification de facteur de confusion. Enfin, on peut se fier pragmatiquement au contrôle de l'erreur de type I quantifiée par le facteur d'inflation génomique (et représentée par le diagramme quantile-quantile). Quelle que soit l'approche choisie, on peut craindre un surajustement et donc une perte de puissance dans le cas où la composante principale ne représente pas un facteur de confusion mais le signal d'association lui-même.

Dans notre cas, l'approche « outlier » aboutirait soit à conserver l'ensemble de l'échantillon (ACP à l'échelle mondiale), soit à l'amputer significativement (ACP à l'échelle locale, par exemple ne garder que les individus aux coordonnées positives sur la composante 1 et négatives sur la composante 2) (Figure 36A). L'application de la statistique de Tracy-Widom aboutit à la sélection de 24 à 39 composantes principales significatives selon le seuil choisi (0.05 ou  $10^{-12}$ , recommandé par (241)), ce qui paraît démesuré. L'inspection de la distribution des valeurs propres (Figure 38A) suggère les seuils possibles de 1, 4 ou 7. Ajuster sur les quatre premières composantes principales aurait l'intérêt conjoint d'ajuster à la fois sur les composantes auxquelles on sait rattacher un sens

(population, effet batch) et celle potentiellement confondantes (la quatrième composante, corrélée au statut cas-témoin).

L'analyse des facteurs d'inflation génomique dans chaque cas (Figure 38B-C) montre un facteur d'inflation d'emblée faible (1.021 sans ajustement) mais diminuant jusqu'à l'inclusion de quatre composantes (1.014). Plusieurs arguments convergent donc vers le choix d'ajuster sur les quatre premières composantes principales. L'alternative recevable à cette approche est de n'ajuster sur aucune composante.

Nous avons sélectionné les 100 meilleurs signaux en l'absence d'ajustement et analysé la variation de leur pvalue après divers ajustements (Figure 38D). Des variations non-négligeables de pvalues après ajustement sont observées. L'ajustement fait perdre jusqu'à 0.75 log à certaines pvalues du top100. Le coefficient de corrélation global des log-pvalues du top100 est de 0.72 seulement pour l'ajustement sur les quatre premières composantes. Cette constatation nous a poussé à conserver les deux approches, sans et avec ajustement sur les quatre premières composantes principales.

## Prise en compte de la structure génétique de la cohorte primaire : ajustement sur composantes principales et modèle mixte

Les diagrammes quantile-quantile (Q-Q plots) et le facteur d'inflation génomique pour le modèle linéaire standard, le modèle linéaire après ajustement sur les quatre premières composantes principales et pour le modèle mixte, sont donnés en Figure 38B,C,E (modèle génétique additif). Le premier constat est que l'approche la plus naïve, un modèle de régression logistique sans ajustement sur les composantes principales, présente un facteur d'inflation génomique de 1.021, ce qui est d'emblée faible donc rassurant. L'ajustement sur les quatre premières composantes principales amène ce facteur d'inflation à 1.014. Le modèle mixte bénéficie du meilleur facteur d'inflation à 1.005.

On peut conclure que dans notre cas les structures génétiques observées n'ont pas agi comme d'importants leviers d'erreur de type I. L'approche par modèle mixte semble marginalement la plus à même de contrôler l'effet des liens familiaux cryptiques et pourrait être préférée à l'ajustement sur composantes principales. Dans notre cas, il semble qu'il n'y ait pas d'argument décisif pour préférer telle ou telle méthode d'analyse. L'observation des « Manhattan plots » révèle cependant que les meilleurs signaux identifiés par chaque approche, susceptibles d'être génotypés dans la cohorte de réplication, ne sont pas les mêmes. Dans ces conditions, préférer un modèle plutôt qu'un autre semble relativement arbitraire et infondé. Nous avons donc choisi de conserver les différents types d'analyse.

La régression logistique sans ajustement reste une méthode de base et bénéficie d'un facteur d'inflation génomique modéré ici ; en l'absence de données pangénomiques au moment de la réplication des meilleurs signaux, elle sera nécessairement adoptée à cette étape. En fonction du nombre de marqueurs génotypés dans la cohorte de réplication, le calcul de composantes principales ou d'une matrice de similarité génétique précise sera en effet difficile voire impossible à l'étape de la réplication. La régression logistique avec ajustement sur les composantes principales est jusqu'ici la

méthode la plus citée pour prendre en compte d'éventuels facteurs de confusion par structure génétique. Le modèle mixte semble le plus adapté à la gestion de liens familiaux cryptiques et bénéficie du meilleur facteur d'inflation génomique.

Par ailleurs, nous avons également utilisé une approche par analyse de survie avec et sans prise en compte des liens familiaux cryptiques. La corrélation entre ces différentes analyses fait donc l'objet de l'étape suivante de notre étude.

		Additive					Dominant					Recessive					Genotypic				Heterozygous				
		GLM	GLM PC	MM	Cox	Cox ME	GLM	GLM PC	MM	Cox	Cox ME	GLM	GLM PC	MM	Cox	Cox ME	GLM	GLM PC	Cox	Cox ME	GLM	GLM PC	MM	Cox	Cox ME
		[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
GLM	[1,]	100	65	88	38	36																			
GLM-PC	[2,]	65	100	66	34	34																			
MM	[3,]	88	66	100	36	34																			
Cox	[4,]	38	34	36	100	93																			
CoxME	[5,]	36	34	34	93	100																			
GLM	[6,]	57	50	55	27	26	100	70	89	40	40														
GLM-PC	[7,]	49	61	47	27	27	70	100	73	39	40														
MM	[8,]	52	49	54	26	25	89	73	100	39	40														
Cox	[9,]	27	28	26	50	47	40	39	39	100	94														
CoxME	[10,]	28	28	27	52	50	40	40	40	94	100														
GLM	[11,]	4	2	4	4	3	0	0	0	0	0	100	86	69	50	52									
GLM-PC	[12,]	3	3	3	4	3	0	0	0	0	0	86	100	65	47	49									
MM	[13,]	4	3	4	4	3	0	0	0	0	0	69	65	100	39	40									
Cox	[14,]	3	2	3	5	4	0	0	0	0	0	50	47	39	100	92									
CoxME	[15,]	2	2	2	5	4	0	0	0	0	0	52	49	40	92	100									
GLM	[16,]	32	27	32	18	18	36	33	36	21	21	34	34	30	22	24	100	71	50	48					
GLM-PC	[17,]	24	30	27	16	16	29	32	31	20	21	30	33	27	18	20	71	100	37	36					
Cox	[18,]	16	14	15	25	24	21	17	20	32	30	21	19	19	25	26	50	37	100	90					
CoxME	[19,]	15	15	14	22	21	20	17	19	32	31	21	19	20	25	27	48	36	90	100					
GLM	[20,]	23	23	21	11	11	43	37	40	23	22	0	0	0	0	0	43	34	30	27	100	82	87	47	46
GLM-PC	[21,]	24	28	22	11	11	40	44	41	22	22	0	0	0	0	0	42	38	30	27	82	100	81	46	44
MM	[22,]	20	19	20	9	9	40	36	43	23	22	0	0	0	0	0	44	36	31	28	87	81	100	48	47
Cox	[23,]	16	15	16	20	19	25	23	26	40	39	0	0	0	0	0	35	27	48	46	47	46	48	100	93
CoxME	[24,]	16	14	16	20	19	25	22	25	40	39	0	0	0	0	0	35	26	48	46	46	44	47	93	100

**Figure 39. Intersections des 100 signaux les plus significatifs pour chaque analyse**

Pour chaque modèle statistique et chaque modèle génétique, les 100 SNPs aux p-values les plus significatives ont été extraits. Cette matrice représente le nombre de SNPs communs parmi par paire de modèle. Sur la diagonale, l'intersection d'un modèle avec lui-même donne 100.

*GLM : modèle linéaire généralisé standard, GLM-PC : modèle linéaire généralisé avec ajustement sur 4 composantes principales ; MM : modèle linéaire mixte incorporant la matrice de similarité entre individus ; Cox : analyse de survie standard ; CoxME : analyse de survie incorporant la matrice de similarité entre individus.*

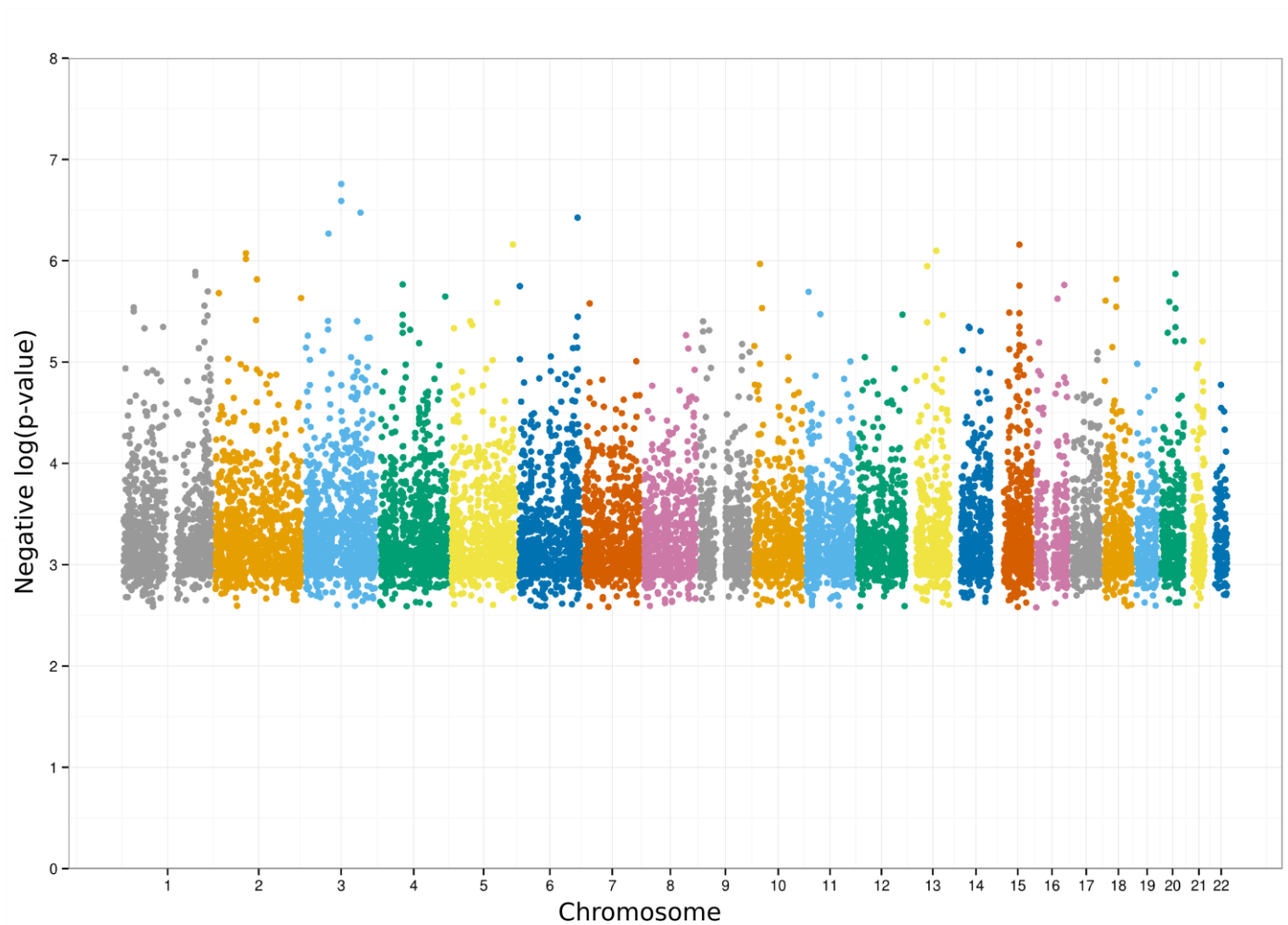
## Corrélation entre analyses

Pour évaluer en pratique la redondance des différentes approches utilisées, nous avons utilisé une approche simple basée sur les 100 meilleurs signaux produits par chaque type d'analyse. L'objectif était de sentir de façon palpable l'impact de la préférence de telle ou telle analyse sur la sélection d'un nombre restreint de SNPs pour génotypage dans la cohorte de réplication. Le nombre de SNPs en commun dans la liste des 100 meilleurs SNPs de chaque paire d'analyse est donné en Figure 39. Autrement dit, il s'agit du cardinal de l'intersection des ensembles des 100 meilleurs signaux de chaque analyse.

L'observation de la matrice met en évidence différents aspects d'intérêt :

- les intersections au sein d'un modèle génétique sont plus grandes qu'au sein d'un modèle statistique. Par exemple, deux analyses avec un codage additif ont plus de similarité que deux analyses utilisant un modèle mixte sous des modèles génétiques différents.
- au sein d'un modèle génétique donné, l'approche binaire (atteint/non atteint) et l'approche par analyse de survie (prise en compte de l'âge de survenue de la maladie) forment deux groupes relativement distincts.
- l'analyse par modèle mixte donne une liste de 100 meilleurs signaux plus proche de la liste des 100 meilleurs signaux du modèle linéaire généralisé sans ajustement que du modèle linéaire généralisé avec ajustement.
- l'analyse de survie avec et sans effet aléatoire donnent des résultats extrêmement proches quel que soit le modèle génétique.
- le modèle génotypique est clairement le plus généraliste, avec la plus grande intersection moyenne avec n'importe quel autre modèle génétique.
- le modèle génétique récessif est le modèle le plus isolé, avec une intersection vide avec les autres modèles, à l'exception du modèle génotypique.





**Figure 40. Manhattan plot des p-values minimales : SNPs génotypés**

Le logarithme négatif de la p-value d'association la plus significative pour chaque SNP parmi 24 tests est représentée en fonction de sa position sur le génome. Seuls les 12091 SNPs les plus significatifs, correspondant à l'union des 2500 meilleurs SNPs de chaque analyse, sont représentés.

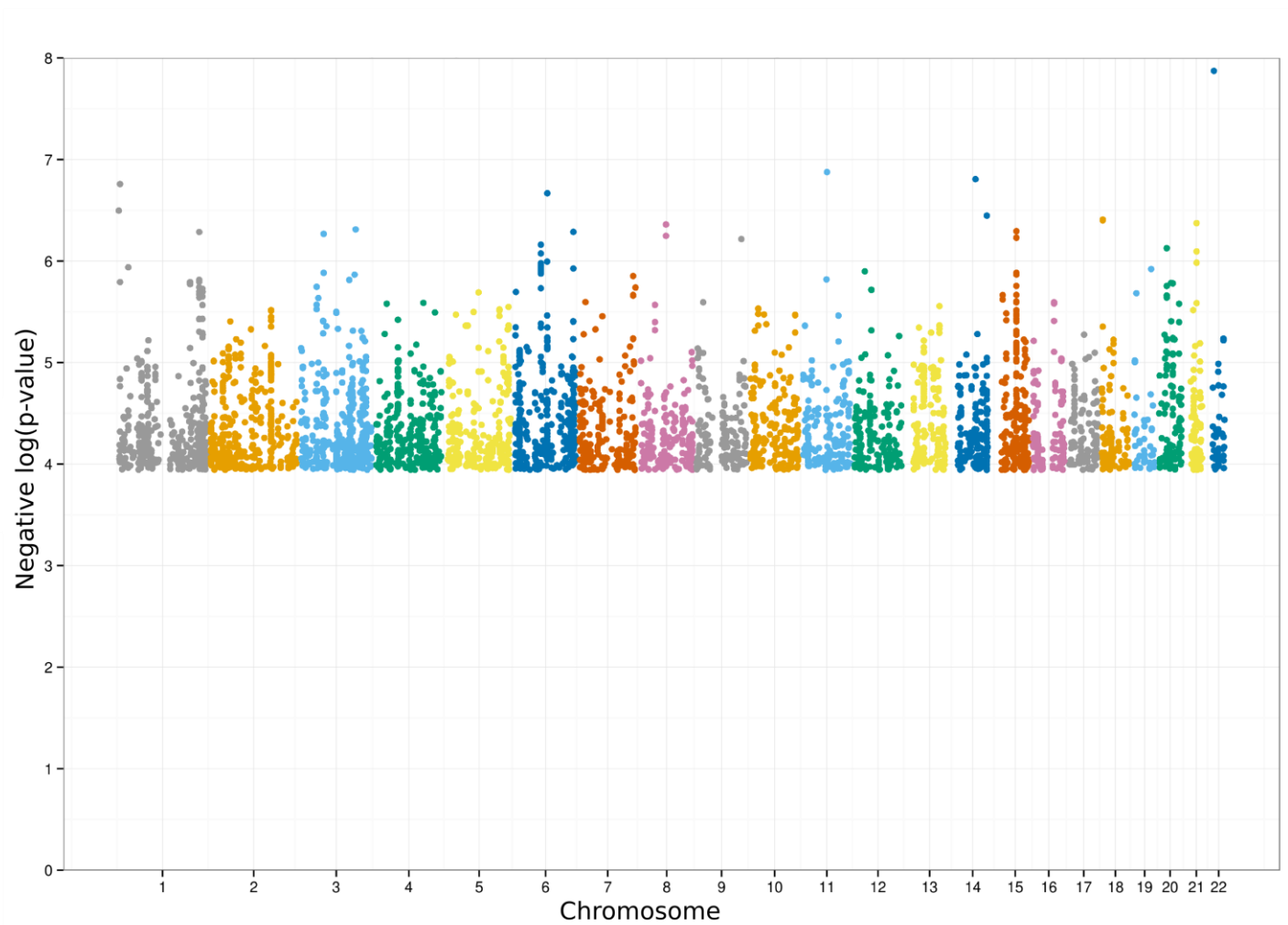
SNP	Chr.	Position (bp)	Gene Name	Transcript Biotype	Consequence to transcript	Regulatory Feature
rs115809482	1	28590183	TAF12	processed_transcript	intron_variant	yes
			RAB42	protein_coding	lessthan5kb_upstream_gene	yes
rs10915181	1	28614821	TAF12	protein_coding	intron_variant	
rs16850799	1	202950723	ADIPOR1	protein_coding	intron_variant	
rs12039275	1	202956762	ADIPOR1	protein_coding	intron_variant	
rs114563230	1	228360415	OBSCN	protein_coding	missense_variant (Polyphen:benign)	yes
rs12144410	1	237891504	ZP4	protein_coding	lessthan5kb_upstream_gene	
			RP11-193H5.1	antisense	intron_variant	
rs116463230	2	9167697			intergenic	
rs13432342	2	85187524	TCF7L1	protein_coding	intron_variant	
			TCF7L1-IT1	sense_intronic	lessthan5kb_downstream_gene	
rs874838	2	85194568	TCF7L1	protein_coding	intron_variant	
rs72828556	2	115734380	DPP10	protein_coding	intron_variant	
rs11890259	2	239942850	MIR4786	miRNA	lessthan5kb_downstream_gene	
			NDUFA10	protein_coding	intron_variant	
rs79475128	3	63546075	RNA5SP134	rRNA	lessthan5kb_downstream_gene	
rs79475128	3	63546075	SYNPR	nonsense_mediated_decay	intron_variant	
			SYNPR-AS1	antisense	intron_variant	
rs4928134	3	99472232			intergenic	yes
rs34060873	3	99474444			intergenic	yes
rs7629237	3	153757368	RP11-23D24.2	antisense	intron_variant	
rs2048751	4	63491796	RP11-12K22.1	lincRNA	intron_variant	
rs10019215	4	183620776			intergenic	
rs10043738	5	129590902	CTC-575N7.1	antisense	intron_variant	
			ADAMTS19	protein_coding	intron_variant	
rs1106693	5	173925760	CPEB4	protein_coding	intron_variant	yes
rs76647377	6	984961			intergenic	
rs9295218	6	163267438	PACRG	protein_coding	intron_variant	
rs10950576	7	15767362			intergenic	
rs1341563	10	16020481			intergenic	
rs16921779	10	21367265	RP11-275N1.1	lincRNA	intron_variant	
rs11041127	11	6992835			intergenic	yes
rs626403	13	50761256	DLEU7	protein_coding	intron_variant	yes
rs9530599	13	76809661			intergenic	yes
rs3743343	15	67194437	SMAD3	protein_coding	3_prime_UTR_variant	
rs1052488	15	67194509	SMAD3	protein_coding	3_prime_UTR_variant	
rs78277697	16	60995693			intergenic	
rs11639802	16	80021988			intergenic	
rs116404259	18	2471151			intergenic	
rs515106	18	35076971	MAPRE2	protein_coding	intron_variant	yes
rs668747	18	35094469	MAPRE2	protein_coding	intron_variant	
rs8115469	20	22371711			intergenic	yes
rs3092030	20	40960888			intergenic	yes
rs1980479	20	40970361			intergenic	yes

Tableau 9. Gènes et transcrits associés aux 37 signaux issus de SNPs géotypés ayant une pvalue inférieure à  $10^{-5.5}$

## Pvalues minimales : Manhattan plot et meilleurs signaux

Dans la section précédente, nous avons vu que les différentes approches d'analyse ne donnent pas la même liste de 100 meilleurs signaux. Or il est difficile de donner une préférence à l'une ou l'autre approche sur des arguments objectifs. Une possibilité pragmatique dès lors est d'assigner à chaque SNP sa pvalue minimale pour les 24 analyses afin d'identifier les signaux les plus prometteurs. Le manhattan plot correspondant à cette stratégie est donné en Figure 40.

37 signaux ont une pvalue de  $10^{-5.5}$  ou moins. La liste des SNPs et les gènes associés dans Ensembl76 sont donnés en Tableau 9.



**Figure 41. Manhattan plot des p-values minimales : SNPs imputés**

Le logarithme négatif de la p-value d'association la plus significative pour chaque SNP parmi 5 modèles génétiques est représentée en fonction de sa position sur le génome.  
Seuls les 5000 SNPs les plus significatifs sont représentés.

SNP	Chr.	Position (bp)	Gene Name	Transcript Biotype	Consequence to transcript	Regulatory Feature
rs4233033	1	2436793	PLCH2	protein_coding	intron_variant	
rs34859510	1	5825202			intergenic	
rs114564953	1	228489725	RNF187	processed_transcript	intron_variant	
rs79475128	3	63546075	RNA5SP134	rRNA	lessthan5kb_downstream_gene	
			SYNPR	protein_coding	intron_variant	
rs79475128	3	63546075	SYNPR-AS1	antisense	nc_transcript_variant	
rs7629237	3	153757368	RP11-23D24.2	antisense	nc_transcript_variant	
rs115891644	6	71637476			intergenic	
rs111745093	6	71659930			intergenic	
rs6937244	6	89758746	MDN1	protein_coding	intron_variant	
rs9295218	6	163267438	PACRG	protein_coding	intron_variant	
rs7821137	8	71245913	EYA1	protein_coding	intron_variant	
rs35975914	8	71246814	EYA1	protein_coding	intron_variant	
rs2797947	9	123376682	DENND1A	protein_coding	lessthan5kb_downstream_gene	
			CRB2	protein_coding	intron_variant	
rs7929389	11	69208515			intergenic	
rs10147945	14	70118301	SLC8A3	protein_coding	intron_variant	
rs1052488	15	67194509	SMAD3	protein_coding	3_prime_UTR_variant	
rs7162810	15	67355352	IQCH	protein_coding	intron_variant	
rs115217119	18	2474604			intergenic	
rs75616506	18	2475149			intergenic	
rs6137616	20	22357742			intergenic	
rs142704870	21	30943953			intergenic	
rs149256730	21	30975231			intergenic	
rs78859	22	20649707	SLC9A3P2	processed_pseudogene	lessthan5kb_upstream_gene	yes
			SMPD4P1	processed_transcript	nc_transcript_variant	yes

Tableau 10. Gènes et transcrits associés aux 23 signaux issus de SNPs imputés ayant une pvalue inférieure à  $10^{-6}$

## Association : cas des SNPs imputés

Les SNPs imputés ont été analysés selon un seul modèle statistique, implémenté dans SNPtest. Il s'agit d'un modèle linéaire généralisé de type régression logistique, prenant en compte le vecteur de probabilité des trois génotypes possibles à chaque SNP pour chaque individu. Ce modèle statistique est appliqué aux cinq modèles génétiques vus précédemment.

Le détail de chaque Manhattan Plot se trouve en annexe. Comme pour les SNPs génotypés, nous avons assigné à chaque SNP sa pvalue minimale. Le Manhattan Plot correspondant est donné en Figure 41.

23 signaux ont une pvalue de  $10^{-6}$  ou moins. La liste des SNPs et les gènes associés dans Ensembl76 sont donnés en Tableau 10.

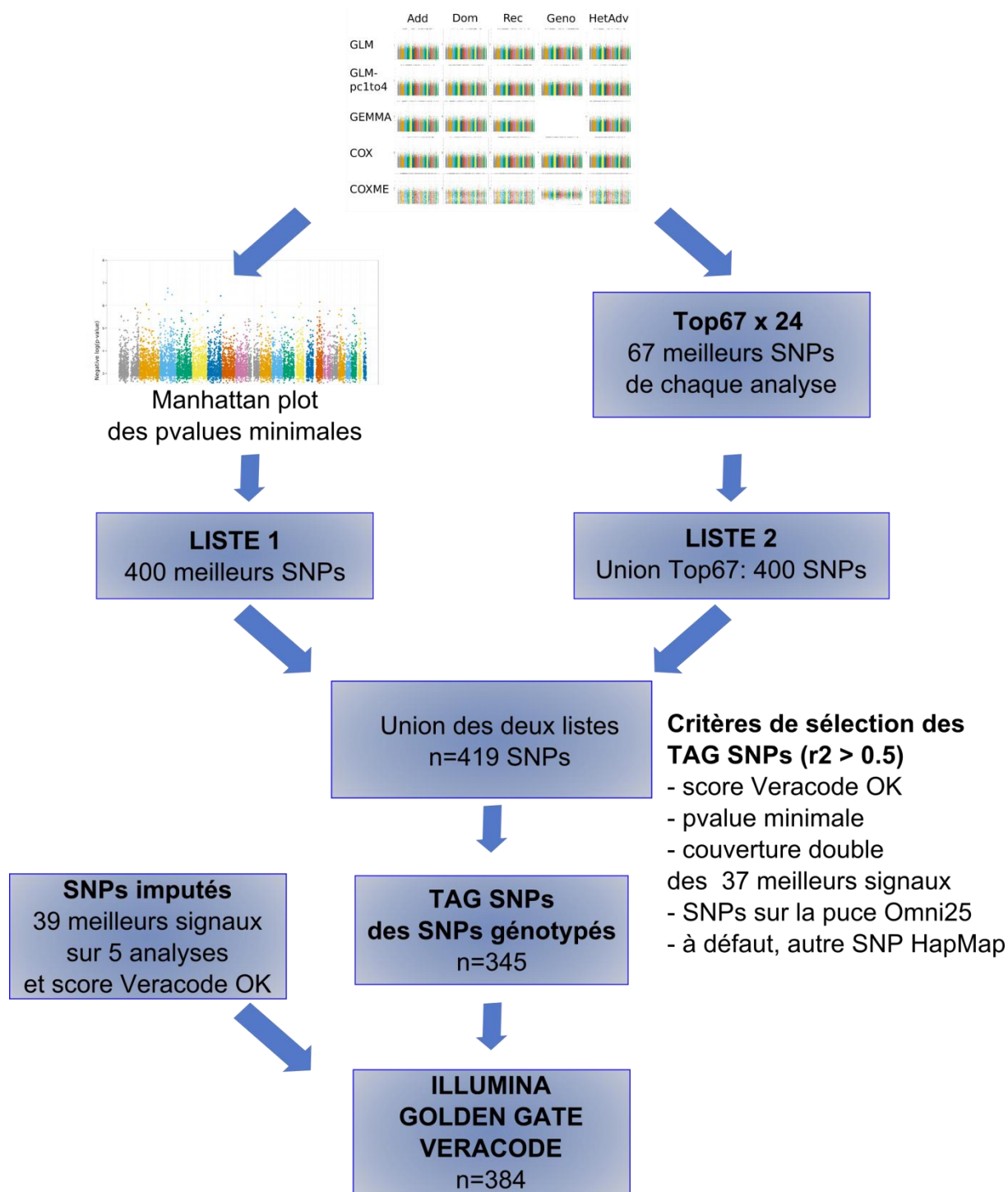


Figure 42. Mode de sélection des 384 SNPs génotypés dans la cohorte de réplication

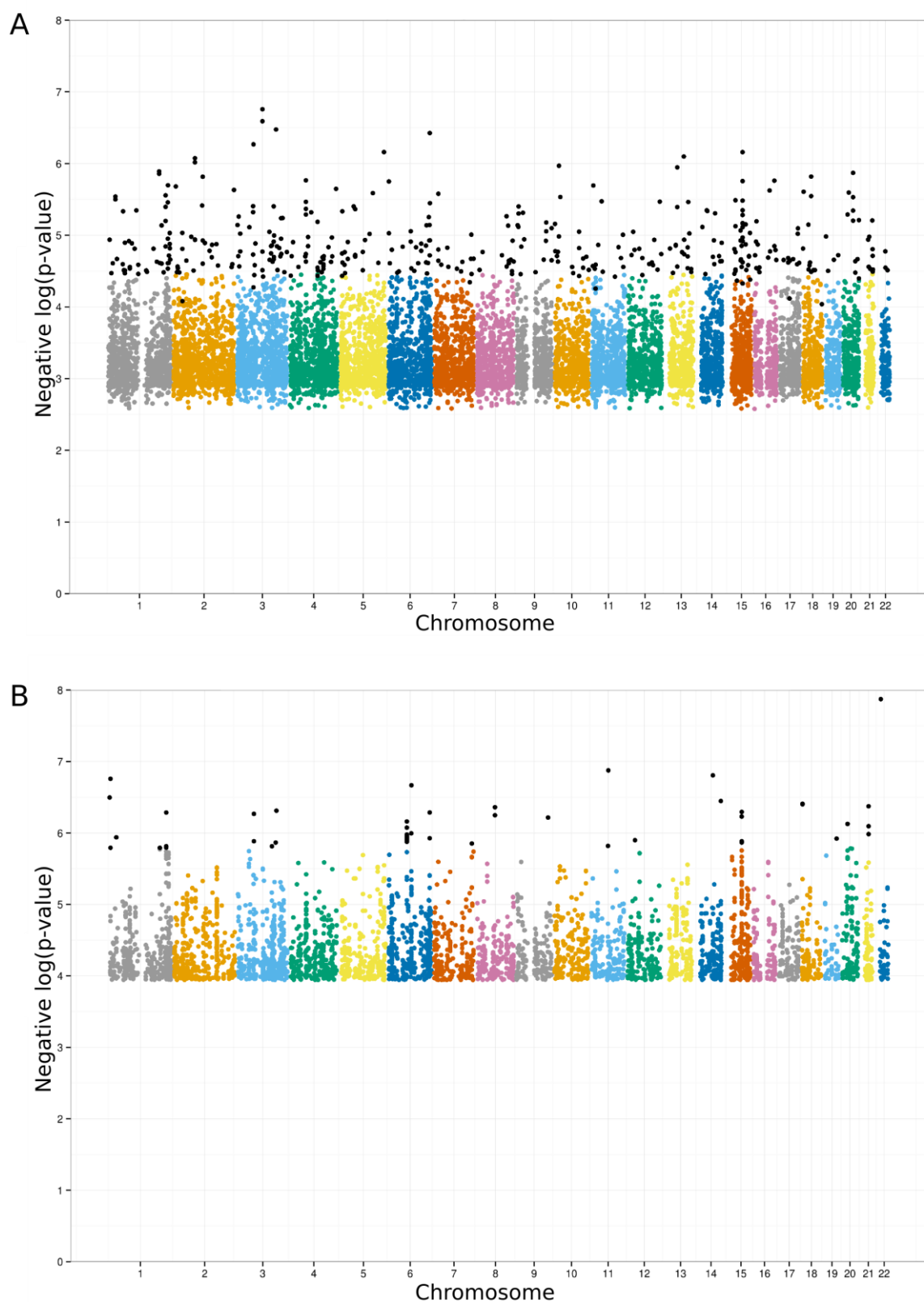
## Réplication des signaux de la cohorte primaire

Au cours de ce travail, nous avons réalisé des millions de tests statistiques. Les analyses d'association pangénomique en général, et la nôtre en particulier, interrogent la problématique des tests multiples avec une intensité inédite dans l'histoire de la théorie statistique. Les corrections faisant l'hypothèse de tests indépendants (Sidak, Bonferroni) semblent trop sévères dans une perspective de génération d'hypothèses, et aboutissent à l'élimination excessive de faux négatifs. La manière la plus raisonnable de se départir des faux positifs tout en contrôlant en partie l'erreur de type II est le passage des meilleurs signaux au révélateur de l'étude de réplication (267). La réplication consiste à tester les meilleurs signaux dans un nouvel échantillon indépendant et de ne retenir que les signaux significatifs à deux reprises.

## Sélection des SNPs pour l'étude de réplication

Nous avons choisi de sélectionner de l'ordre de 400 SNPs pour génotypage dans la cohorte de réplication (Figure 42). Pour sélectionner ces SNPs, une première possibilité est de prendre les 400 meilleurs SNPs de l'approche pvalue minimale. Une autre stratégie possible est une approche paritaire, c'est-à-dire la sélection des n meilleurs signaux de chacune des 24 analyses. Dans notre cas, la liste de 67 meilleurs signaux de chacune des 24 analyses aboutit à une liste de 400 SNPs, du fait de la présence de nombreux SNPs dans plusieurs listes. L'union de la liste des 400 meilleurs SNPs de l'approche pvalue minimale et des 400 SNPs issus de l'union des 67 meilleurs SNPs de chaque analyse aboutit à une liste de 419 SNPs. Les deux approches sont donc largement redondantes. Les SNPs sélectionnés sont représentés en noir sur la Figure 43A.





**Figure 43. Manhattan plot : SNPs sélectionnés pour la réplication**

- (A) Les 419 SNPs génotypés sélectionnés pour la réplication sont mis en évidence en noir.
- (B) Les 60 SNPs imputés sélectionnés pour la réplication sont mis en évidence en noir.

Cette liste de 419 SNPs est constituée en partie de SNPs non indépendants. Le déséquilibre de liaison à certains loci rend redondant l'information apportée par plusieurs SNPs adjacents. Il est classique de réduire la liste des SNPs à une liste de SNPs indépendants, ce qui permet finalement de tester plus d'hypothèses indépendantes d'association génétique. Nous avons utilisé le logiciel Haploview, le logiciel R et l'interface en ligne SNAP pour sélectionner une liste de signaux indépendants (172,268,269).

Le génotypage a été réalisé sur une puce Illumina à façon de technologie Golden Gate Veracode. Un critère majeur de sélection des SNPs pour la réplication est le score de faisabilité biochimique assigné par Illumina pour chaque SNP. Ce score est fonction de la séquence flanquante du SNP, de sa spécificité, et de la présence d'autres SNPs très proximaux.

Les critères utilisés pour sélectionner les SNPs ont donc été les suivants : à chaque locus, est conservé le SNP le plus significatif passant le score biochimique Illumina. Les SNPs en déséquilibre de liaison avec un coefficient de détermination  $r^2$  supérieur à 0.5 sont éliminés. Les 37 meilleurs signaux sont en revanche sécurisés par 2 SNPs. Pour les signaux isolés concernant des SNPs ayant un score biochimique insatisfaisant, un SNP en  $r^2$  supérieur à 0.5 dans la population HapMap Yoruba est sélectionné.

Cette procédure a permis la réduction de la liste initiale de 419 SNPs à 345 SNPs pour la réplication. Nous avons par ailleurs sélectionné les 39 meilleurs signaux issus de l'analyse après imputation pour compléter la liste à 384 SNPs (en noir sur la Figure 43B).

## Cohorte de réplication

La cohorte de réplication est un échantillon indépendant de cas et de témoins recrutés selon les mêmes critères et au même endroit que la cohorte primaire. Nous avons reçu très récemment le génotypage de cette cohorte de 708 individus, dont 693 passent le contrôle qualité par individu. 4 individus sont éliminés pour un taux de génotypage inférieur à 95% et 11 échantillons sont éliminés pour duplication (distance IBS < 1%). Ces 693 individus sont répartis en 454 cas et 239 témoins. La proportion de femmes parmi les cas et témoins était respectivement de 56 et 60%. L'âge médian des cas était de 15 ans et celui des témoins de 23 ans. Les cas sont répartis en deux catégories : 230 sont des cas confirmés par PCR, et 224 sont des cas pour lesquels aucun examen de laboratoire n'a été retrouvé (n=180) ou un examen négatif avec une forte suspicion clinique (n=44).

Les résultats de l'analyse de réplication seront présentés lors de la soutenance orale le 28 novembre.

## Etude d'association pangénomique : discussion

Nous rapportons la première étude pangénomique de l'architecture génétique de la susceptibilité à l'ulcère de Buruli. La taille de notre échantillon, modeste dans le contexte général des GWAS, est remarquable dans le contexte d'une maladie infectieuse tropicale négligée comme l'UB. La qualité diagnostique est également remarquable, puisque la cohorte primaire s'appuie intégralement sur des cas d'UB confirmés par PCR.

Nous avons fait face aux difficultés inhérentes à la réalisation d'un projet scientifique à dimension technologique poussée dans une zone rurale reculée du Bénin. L'adhésion de la population de Pobè et alentours à un sujet aussi étrange et étranger que l'exploration de facteurs génétiques impliqués dans le développement de l'UB a été le reflet d'une grande confiance instaurée année après année par l'équipe du CDTUB. Sur le plan technique, notre étude démontre la faisabilité d'une telle recherche dans des conditions tropicales. Le catalogue des GWAS du NHGRI est encore largement dominé par l'étude de populations d'origine européennes (223). Le succès technique de notre étude, avec plus de 99% d'échantillons dont le taux de génotypage dépasse 95%, doit rassurer les investigateurs.

L'histoire démographique africaine pose ses difficultés propres en termes d'association génétique. L'ancienneté des populations africaines, la grande diversité génétique et le niveau modéré de déséquilibre de liaison posent des défis stimulants pour la recherche en génétique humaine (270). Dans l'attente de l'analyse des données de réplication, il est difficile de donner des conclusions définitives sur les signaux identifiés. Il est intéressant de noter que l'un des principaux signaux d'association de la susceptibilité à la lèpre, impliquant le gène *PACRG*, est présent dans notre liste de 37 signaux majeurs (161). De nombreux autres signaux ont un intérêt biologique, dont *SMAD3* impliqué dans les processus de cicatrisation. Par ailleurs, il faudra prendre en compte la

structure de déséquilibre de liaison autour des signaux répliqués pour identifier le SNP vraisemblablement causal.

Les perspectives de ce travail sont nombreuses. L'analyse du chromosome X, non finalisée, n'a pas été présentée ici. La présence diploïde de ce chromosome chez la femme et haploïde chez l'homme impose des méthodes d'analyse particulières (271). Par ailleurs, l'inactivation de certaines régions du chromosome X modifient également l'approche standard d'analyse. L'analyse des cascades de signalisation (« pathway ») est également prometteuse. Cette approche, qui introduit certaines connaissances biologiques *a priori* dans le modèle d'analyse, permet de rechercher si les signaux d'association sont regroupés de manière statistiquement significative dans certaines voies de signalisation (272,273). Enfin, nos analyses ont essentiellement porté sur les SNPs génotypés. Les SNPs imputés n'ont pas bénéficié de l'exploration par modèle mixte par exemple, ni d'analyse de survie. Au-delà des variations ponctuelles, l'analyse des variations en nombre de copies (CNV) est également une perspective importante.

Les implications cliniques et biologiques de l'étude de la génétique humaine de l'infection par *M. ulcerans* sont multiples. Sur le plan biologique, il s'agit de permettre une compréhension de la physiopathologie de l'infection, et d'évaluer la fonction des gènes dans un environnement naturel. En cela, la génétique humaine est le complément de la génétique de l'animal, chez lequel le processus infectieux est parfaitement contrôlé mais modélisé selon un protocole éloigné de l'histoire naturelle de l'infection, en terme d'inoculum et de mode d'exposition. Sur le plan clinique, les implications, moins immédiates, sont potentiellement majeures : contribuer au diagnostic étiologique de la maladie infectieuse, l'identification du germe étant complétée par l'identification de la prédisposition génétique; développer de nouvelles thérapeutiques, visant à compléter le défaut immunitaire de l'hôte plutôt qu'à s'attaquer directement à l'agent infectieux; augmenter la puissance des approches de prévention, en cherchant une efficacité vaccinale chez les individus susceptibles, plutôt que dans une population composée en partie d'individus naturellement

résistants à l'infection. La génétique de l'animal est une génétique *in vivo*, souple mais imparfaitement physiologique. La génétique humaine, qui explore la diversité génétique telle qu'elle existe, est une génétique *in natura*, complément nécessaire aux organismes modèles, puisqu'elle seule observe l'infection et l'immunité dans leur écosystème naturel (274).

J'aimerais orienter la suite de cette discussion sur l'approche GWAS en elle-même et les perspectives d'évolution plus générales de la génétique humaine des maladies complexes. Les études d'association pangénomique ont connu une grande popularité à partir de 2005. Ces études ont été un moteur pour la mise en place de collaborations impliquant en général de nombreuses équipes internationales et rassemblant parfois un nombre astronomique de patients et de témoins (plusieurs milliers à dizaines de milliers). Les méta-analyses de GWAS atteignent des sommets étonnants, avec jusqu'à 260 000 individus dans une récente étude sur la susceptibilité génétique à l'obésité (275). La dimension de ces études semble justifier parfois à lui seul leur publication dans des revues de très haut niveau. Sur les 1961 publications recensées en août 2014 par le catalogue des GWAS (certains articles comportant plusieurs phénotypes), 389 ont été publiés dans le premier journal de spécialité génétique, *Nature Genetics*, et 459 au total dans les revues à facteur d'impact supérieur à 30 (*Nature*, *Science*, *New England Journal of Medicine*, *Lancet*, *Nature Genetics*), soit près de 25% (223). Dans le domaine des maladies infectieuses, une centaine de phénotypes ont été explorés, dont 18 phénotypes concernant l'infection par le VIH, 15 phénotypes concernant le virus de l'hépatite C, 10 phénotypes concernant le virus de l'hépatite B, 8 phénotypes concernant la tuberculose et 4 le paludisme. Les populations explorées par GWAS sont très majoritairement caucasiennes.

Malgré ce succès éditorial, plusieurs voix se sont élevées pour critiquer les GWAS. Ces critiques sont de plusieurs ordres : certains estiment que l'allocation de ressources économiques pour les GWAS a été disproportionnée, d'autres estiment que les résultats scientifiques ont été décevants et n'ont pas contribué à la compréhension physiopathologique des maladies, et certains cliniciens (particulièrement pressés) estiment que le bénéfice des GWAS pour le patient est

inexistant (276). Une frange de la communauté scientifique de la génétique épidémiologique même s'est levée contre le GWAS qui, étonnamment, parvient à faire de la génétique sans échantillonner de familles. Une tension est également née du caractère presque automatisé en apparence de la production de GWAS par des consortiums de plus en plus grands. Le GWAS a probablement transposé à la génétique humaine une opposition entre « big science » et « small science », dont les objectifs prioritaires sont parfois divergents, puisqu'il est difficile d'atteindre certaines tailles d'échantillon si l'on s'astreint à une certaine qualité et homogénéité dans la description du phénotype d'intérêt. Quoi qu'il en soit, on peut rester quelque peu dubitatif face à un article rapportant pas moins de 7 GWAS sur des phénotypes indépendants dans un seul et même texte (259). C'est l'aspect mécanique du GWAS qui prévaut alors, puisqu'il mêle sur quelques pages des résultats concernant des milliers de patients atteints de troubles bipolaires, des milliers de patients atteints de diabète de type II et des milliers de patients ayant fait un infarctus du myocarde.

Un paramètre célèbre de la génétique quantitative animale a pris une place de choix dans le débat sur l'efficacité ou l'inefficacité des GWAS en génétique humaine: l'héritabilité. Les GWAS seraient inefficaces puisque les variants identifiés n'expliquent qu'une proportion modeste (souvent moins de 10%, au mieux 20 à 30%) de l'héritabilité estimée d'une maladie. Le papier ayant introduit le concept d'héritabilité manquante, post-GWAS, a été cité plus de 3000 fois en 5 ans (277). La notion d'héritabilité manquante fait consensuellement partie de toute introduction de demande de financement, ou de publication, d'une étude génétique sur une maladie complexe. Je me sens personnellement mal à l'aise avec ce concept, dont l'utilisation pourrait faire preuve d'un certain opportunisme. Les limitations dans l'estimation de l'héritabilité sont en effet tellement manifestes dans le cadre de la génétique humaine que, pour pouvoir poursuivre sa recherche, chacun s'empresserait de les rappeler si cette estimation aboutissait à l'absence d'héritabilité manquante.

L'héritabilité voudrait mesurer la proportion de la variance phénotypique d'un trait ou d'une maladie au sein d'une population qui soit attribuable à des facteurs génétiques. C'est un concept

développé d'abord pour l'analyse de traits quantitatifs, puis étendue aux traits binaires comme les maladies par l'introduction d'une échelle de risque sous-jacente et nécessitant une estimation de la prévalence de la maladie. Il y a probablement plus d'articles sur les erreurs d'interprétation du concept d'héritabilité que sur l'héritabilité elle-même. Les titres sont éloquents : « Heritability, one word, three concepts » ; « Heritability in the genomics era – concepts and misconceptions » ; « The heritability of human disease : estimation, uses and abuses » (278–281). La controverse sur l'interprétation de l'héritabilité en génétique humaine a commencé il y a plus de 40 ans et se poursuit toujours (282,283). Les calculs de l'héritabilité sont essentiellement une transformation d'estimations de corrélations phénotypiques entre apparentés, pondérée par les différents degrés d'apparentement. Les hypothèses d'estimation sont extrêmement restrictives : les facteurs génétiques sous-jacents à la maladie seraient indépendants, n'interagissant pas ; leurs effets seraient additifs et ne dépendraient pas de l'environnement du sujet. L'impact de l'environnement sur un phénotype complexe a pu être minimisé en génétique animale et végétale, domaines dans lesquels le concept d'héritabilité a permis des progrès notables dans les processus de sélection artificielle d'individus pour la maximisation de paramètres quantitatifs (masse de graine ou de lait produite par individu) (284). En génétique humaine, les estimations sont sujettes à caution, en particulier pour un trait binaire.

Reprocher aux GWAS d'être inefficaces car ils n'expliquent qu'une faible part de l'héritabilité, c'est considérer l'estimation de l'héritabilité comme correcte et donc considérer que les maladies complexes humaines sont issues de variants à effet additif, sans interaction gène-gène ni interaction gène-environnement. C'est exactement sous ces hypothèses que sont également construits les modèles prédictifs génétiques tentant de quantifier le risque de développement d'une maladie complexe à partir des variants associés en GWAS. Il n'est donc pas particulièrement étonnant de constater leur relative inefficacité actuelle (285–287). Il semble encore actuellement que la prédiction de risque de diabète soit plus fiable par l'interrogatoire simple du patient sur ses



antécédents familiaux et personnels que par le génotypage des variants associés par GWAS (288). Il me semble que c'est plutôt l'architecture du risque génétique qui nous échappe.

La vague de mécontentement anti-GWAS a été suffisamment forte pour générer une série d'articles de revue en défense du GWAS (276,289,290). Ces articles rappellent que les GWAS ont apporté des preuves solides qu'une susceptibilité génétique complexe existe dans de nombreuses maladies communes. En effet, des *milliers* de signaux répliqués ont été listés en quelques années à peine dans le catalogue des GWAS du NHGRI. Ils rappellent également que l'approche agnostique pangénomique du GWAS est une force, puisqu'elle permet de générer des hypothèses de recherche totalement inattendues, comme cela a été le cas par exemple dans la dégénérescence maculaire liée à l'âge, la maladie de Crohn et l'infection par le virus de l'hépatite C. On peut noter qu'il existe quelques exemples d'élucidation fonctionnelle du mécanisme par lequel l'allèle à risque confère ce risque (291,292). Ces mécanismes d'élucidation reposent en général sur l'identification d'un effet du génotype à risque sur le niveau d'expression d'un gène d'intérêt (eQTL), ou de méthylation d'une région donnée (metQTL).

Une difficulté importante dans l'acceptation des GWAS comme ayant contribué aux connaissances scientifiques pourrait venir de la nature de leur contribution : la mise au jour de corrélations, qui questionnent la notion de causalité et de preuve scientifique. L'épidémiologie, certes génétique, s'invite dans des revues scientifiques habituées à publier des résultats de biologie moléculaire. Un article de génétique moléculaire contient fréquemment toute la chaîne causale reliant une mutation à une maladie génétique : la mutation est séquencée, la protéine change de taille, sa fonction est annulée, la complémentation par un allèle sauvage restaure cette fonction. Mais voilà qu'à la place de gels et d'imageries confocales apparaissent des Manhattan plots et tableaux d'odds-ratios dont le maximum dépasse exceptionnellement 2. Les seuils de significativité apparaissent eux-mêmes variables. Nous quittons le cadre de la causalité déterministe : ici, les cas comme les témoins portent, ou ne portent pas, le variant déclaré à risque. Par ailleurs, probablement

pour des questions de finesse de résolution, la biologie moléculaire actuelle peine à identifier des phénotypes cellulaires covariant avec le génotype à risque dans les maladies complexes. Autrement dit, le signal reste statistique. Aucune cause nécessaire ou suffisante n'est dévoilée : une suite de corrélations ni nécessaires ni suffisantes est listée.

Pour aller un peu plus loin, il semble que la génétique statistique souffre d'un mal très mendélien : la volonté de binariser la découverte. Un variant est répliqué ou ne l'est pas. Mais la difficulté à définir ce qu'est un variant répliqué montre d'emblée l'ampleur du problème. La question du seuil de pvalue, par essence arbitraire, est sans cesse reposée. Les variants répliqués publiés sont souvent tout simplement les variants les *mieux* répliqués. Dans cette perspective, il ne s'agit pas tant de valeurs absolues de pvalues que de classement des meilleures pvalues, comme le montre le processus de sélection des SNPs pour la réplication. On risque alors de tomber dans cette idée irritante qu'un GWAS ne perd jamais : à un certain seuil, certains SNPs seront associés. Mais cette formulation elle-même est binaire, comme l'est l'idée que la recherche est une guerre que l'on perd ou l'on gagne. En réalité, si les phénotypes et les génotypes sont correctement caractérisés, un GWAS est ce qu'il est : une somme de corrélations, permettant peut-être de prioriser la suite de la recherche. Voilà peut-être la conclusion : les meilleurs signaux de GWAS sont enrichis en découvertes scientifiques. Ce sont des briques pour le futur.

Les GWAS poussent hors de ses limites la théorie statistique en posant des questions insolubles de seuils de significativité. Je pense également que les GWAS poussent l'esprit humain à ses limites. Comment conceptualiser les résultats d'un GWAS ? Comment se les représenter ? Les GWAS communiquent essentiellement leurs résultats par un tableau de quelques signaux principaux associés au phénotype. Des millions de marqueurs ont été testés, et une vingtaine apparaissent dans la publication et sont commentés. Pourtant, certaines études montrent que la capacité de discrimination des scores de prédiction de risque génétique augmente si l'on y inclut des SNPs faiblement associés sur le critère de la pvalue (287). Ces SNPs n'ont bien sûr pas été génotypés dans

la cohorte de réplication. L'information contenue dans les GWAS dépasse donc largement la vingtaine de SNPs icônisés qui seront cités dans l'abstract ou peut-être même dans le titre. L'auteur même d'un GWAS est généralement incapable de résumer les meilleurs signaux de son étude au-delà peut-être de quelques marqueurs. Il ne peut en faire mémoire ni en dessiner un schéma, puisqu'ils ne sont généralement rattachés à aucun sens unifié et restent potentiellement des listes de marqueurs nombreux et anonymes.

Pour finir, il me semble important de rappeler que les difficultés statistiques posées par les GWAS sont multipliées de façon littéralement exponentielle dès lors qu'on complexifie légèrement le modèle causal envisagé. Pour sophistiqués et immenses qu'ils soient, les GWAS n'en testent pas moins des millions d'hypothèses simples et ouvertement trop simples : l'association marginale d'un allèle à un phénotype clinique. La génétique complexe, qui repose sur l'hypothèse d'une hérédité polygénique impliquant l'interaction de dizaines ou centaines de variants modifiant le risque de la maladie, est essentiellement cantonnée à tester séquentiellement l'effet marginal de chaque variant dans des modèles linéaires univariés. Il semble pourtant extrêmement vraisemblable que certains variants agissent de concert, par exemple dans une voie métabolique. On pourrait imaginer à l'instar de Kenneth Rothman en épidémiologie non génétique que certains complexes causaux, formés d'éléments causaux individuellement ni nécessaires ni suffisants, se révèlent eux-mêmes causes suffisantes (293). Autrement dit, aucun des dix allèles associés à la maladie dans dix gènes n'est suffisant pour déterminer la survenue de la maladie, mais la présence conjointe chez un individu de, disons, 4 de ces variants suffit à la provoquer. Cette théorie séduisante soulagerait le chercheur attaché à la causalité déterministe. Intuitivement, une autre possibilité est que la génétique humaine ait à faire un saut conceptuel en termes de déterminisme causal analogue à celui que la physique mécaniste a fait au cours du XX<sup>ème</sup> siècle pour appréhender une réalité quantique.

Sur le plan des phénotypes d'intérêt, une perspective intéressante de la génétique humaine me semble être la génétique de sous-phénotypes quantitatifs : étudier la glycémie plutôt que le

diabète, ou la production d'une cytokine en réponse à un agent infectieux plutôt que la maladie infectieuse. L'écart et le nombre de facteurs confondants séparant un sous-phénotype d'un allèle modificateur me paraît moins abyssal que celui entre un phénotype clinique et ce même allèle. Ce type d'étude a d'ores et déjà commencé (294). Dans ce contexte, des phénotypes fortement corrélés sont souvent produits (production de telle cytokine en réponse au stimulus A ou B) et l'analyse combinée de ces phénotypes est également une perspective majeure de développement théorique et appliqué.

Sur le plan des génotypes d'intérêt, la génétique humaine complexe va progressivement basculer vers la prochaine révolution technologique : le séquençage du génome entier de larges échantillons de cas et de témoins. Ceci va résoudre deux limites du GWAS : l'identification de variants rares (fréquence de l'allèle mineur inférieur à 1%) impliqués dans la susceptibilité aux maladies complexes et l'incomplète couverture du génome pour les variants communs. Une étude récente montre que les couvertures du génome par les puces à ADN, c'est-à-dire la proportion de variants connus avec une fréquence de l'allèle mineur supérieure à 1% étant en  $r^2 > 0.8$  avec un SNP présent sur la puce, était largement surestimée dans les brochures commerciales des sociétés Affymetrix et Illumina. Dans le meilleur des cas (populations d'origine européenne), cette étude révèle que seuls 50% des variants sont effectivement couverts selon cette définition (295). La meilleure couverture future du génome, associée à l'amélioration progressive de la compréhension du potentiel régulateur des différentes régions du génome humain par l'étude de ses modifications épigénétiques, permettra de faire des pas supplémentaires sur le chemin de ceux qui, en explorant des corrélations, espèrent faire des découvertes.



# Conclusion

---



L'ulcère de Buruli, infection à *M. ulcerans*, est la troisième mycobactériose mondiale. Identifiée en 1897, cette pathologie connaît une émergence rapide depuis 1980, essentiellement dans les pays d'Afrique subsaharienne. Nous avons constitué la plus grande cohorte de cas confirmés à ce jour afin de décrire l'épidémiologie classique de la maladie et d'explorer l'architecture génétique de la susceptibilité à cette maladie.

Les patients atteints d'UB sont des enfants (âge médian au diagnostic de 12 ans), présentant une lésion unique (96%), large (plus de 15 cm, 36%), ulcéral (66%) du membre inférieur (60%). Nous rapportons une présentation clinique atypique de l'UB, dans laquelle les patients présentent une ostéomyélite à *M. ulcerans* sans lésion cutanée active ou passée. Le sex-ratio varie significativement avec l'âge : les garçons sont majoritaires parmi les enfants (57% de patients masculins chez les moins de 15 ans), et les femmes parmi les adultes (33% de patients masculins). La présentation clinique dépend significativement de l'âge et du sexe. En particulier, 9% des patients masculins ont présenté une ostéomyélite contre 4% des patients féminins. Un an après la fin du traitement, 22% des patients présentent des séquelles fonctionnelles fixées. Une présentation clinique comportant une lésion oedémateuse, osseuse, large ou plusieurs lésions est significativement associée avec le développement de séquelles fonctionnelles (OR 7.64, IC95% [5.29-11.31]), ce qui conduit à une définition opérationnelle de la sévérité de l'UB. Les patients coinfectés par le VIH ont un risque significativement plus élevé de développer un UB sévère (OR 2.77, IC95% [1.32-6.33]). Nos connaissances dans l'épidémiologie clinique de l'UB sont encore incomplètes. La dynamique de la guérison de l'UB pourrait bénéficier d'études complémentaires, afin de mieux différencier rechutes infectieuses, rechutes immunologiques et réinfections. Plus généralement, on peut rappeler ici que le mode de transmission de la maladie reste incertain.

Nous avons exploré l'architecture génétique de l'UB dans une perspective mendélienne et une perspective complexe. Le cas le plus sévère de la maladie observé au CDTUB de Pobè appartient à une famille consanguine dans laquelle la ségrégation du phénotype suggère un défaut génétique



mendélien récessif. Une analyse de liaison génétique par cartographie d'homozygotie a permis d'impliquer le locus des bêta-défensines sur le chromosome 8 dans la pathogénèse de l'UB, et d'identifier une délétion homozygote ségrégeant parfaitement avec la maladie. Les perspectives de ce travail sont nombreuses. La caractérisation moléculaire de la délétion et l'expression normale ou altérée des défensines dans les cellules des patientes pourraient faire l'objet d'explorations approfondies. La singulière complexité de cette région répétée du génome fait néanmoins de ce projet de recherche un projet à haut risque. L'identification d'autres familles contribuant positivement à la liaison dans la même région du génome conforterait notablement l'hypothèse générée ici.

Dans une perspective complexe, une étude d'association pangénomique a été réalisée après génotypage d'une cohorte de 400 cas et 400 témoins exposés sur plus de 2 millions de SNPs par la puce Illumina Omni2.5 et a permis l'identification de nombreux signaux d'intérêt. La réalisation de millions de tests statistiques implique une certaine contamination de ces signaux par de faux positifs. C'est l'objectif de l'étude de réplication en cours que de faire la différence entre signal et bruit statistique. Elle sera présentée lors de ma soutenance orale. Cette étude d'association pangénomique reste encore incomplète. L'analyse du chromosome X reste en particulier à réaliser. Par ailleurs, la couverture du génome africain par les puces à ADN n'est pas encore optimale en l'état actuel des technologies. La prochaine révolution en génétique sera celle du séquençage du génome entier de centaines de cas et de témoins. Ces techniques existent d'ores et déjà mais sont à un prix prohibitif pour ces tailles d'échantillons. A couverture complète du génome, de nombreux aspects resteront à explorer, en premier lieu l'immense domaine des effets conjoints de variants génétiques à différents loci (interactions).

La compréhension de la physiopathologie de l'infection à *M. ulcerans* est cruciale pour générer de nouvelles pistes thérapeutiques et vaccinales. La dissection du contrôle génétique de l'infection par l'hôte est en ce sens indispensable.





# Bibliographie

---

1. Davies JNP, Elmes S, Hutt MSR, Mtimavalye LAR, Owor R, Shaper L. Cancer in an African Community, 1897-1956. *BMJ*. 1964 Feb;1(5378):259–64.
2. Billington WR. Albert Cook 1870-1951: Uganda pioneer. *BMJ*. 1970 Dec;4(5737):738–40.
3. Buruli ulcer. *BMJ*. 1970 May;2(5706):378–9.
4. MacCALLUM P, TOLHURST JC. A new mycobacterial infection in man. *J Pathol Bacteriol*. 1948 Jan;60(1):93–122.
5. Van OYE E, BALLION M. [Is it necessary to take into account a new infection from acid-resistant bacilli in Africa? Preliminary note]. *Ann Société Belge Médecine Trop*. 1950 Sep;30(3):619–27.
6. JANSSENS PG, QUERTINMONT MJ, SIENIAWSKI J, GATTI F. Necrotic tropical ulcers and mycobacterial causative agents. *Trop Geogr Med*. 1959 Dec;11:293–312.
7. Clancey J. MYCOBACTERIAL SKIN ULCERS IN UGANDA. *The Lancet*. 1961 Oct;278(7209):951–4.
8. DODGE OG, LUNN HF. Buruli ulcer: a mycobacterial skin ulcer in a Uganda child. *J Trop Med Hyg*. 1962 Jun;65:139–42.
9. Connor DH, Lunn HF. Mycobacterium ulcerans infection (with comments on pathogenesis). *Int J Lepr*. 1965 Sep;33(3):Suppl:698–709.
10. George KM, Chatterjee D, Gunawardana G, Welty D, Hayman J, Lee R, et al. Mycolactone: a polyketide toxin from Mycobacterium ulcerans required for virulence. *Science*. 1999 Feb 5;283(5403):854–7.
11. WHO | The Yamoussoukro Declaration on Buruli ulcer [Internet]. [cited 2011 Apr 25]. Available from: [http://www.who.int/buruli/yamoussoukro\\_declaration/en/index.html](http://www.who.int/buruli/yamoussoukro_declaration/en/index.html)
12. WHO | Global Buruli Ulcer Initiative events [Internet]. [cited 2011 Apr 25]. Available from: <http://www.who.int/buruli/events/en/>
13. WHO. Buruli ulcer epidemiological situation [Internet]. [cited 2014 May 21]. Available from: [http://apps.who.int/neglected\\_diseases/ntddata/buruli/buruli.html](http://apps.who.int/neglected_diseases/ntddata/buruli/buruli.html)
14. Crompton D, World Health Organization. Working to overcome the global impact of neglected tropical diseases : first WHO report on neglected tropical diseases. Geneva Switzerland: World Health Organization; 2010.
15. Amofah G, Bonsu F, Tetteh C, Okrah J, Asamoah K, Asiedu K, et al. Buruli ulcer in Ghana: results of a national case search. *Emerg Infect Dis*. 2002 Feb;8(2):167–70.
16. Amofah GK, Sagoe-Moses C, Adjei-Acquah C, Frimpong EH. Epidemiology of Buruli ulcer in Amansie West district, Ghana. *Trans R Soc Trop Med Hyg*. 1993 Dec;87(6):644–5.
17. Yeboah-Manu D, Röltgen K, Opere W, Asan-Ampah K, Quenin-Fosu K, Asante-Poku A, et al. Sero-epidemiology as a tool to screen populations for exposure to Mycobacterium ulcerans. *PLoS Negl Trop Dis*. 2012 Jan;6(1):e1460.
18. Debacker M, Aguiar J, Steunou C, Zinsou C, Meyers WM, Guédénou A, et al. Mycobacterium ulcerans disease (Buruli ulcer) in rural hospital, Southern Benin, 1997-2001. *Emerg Infect Dis*. 2004 Aug;10(8):1391–8.

19. Porten K, Sailor K, Comte E, Njikap A, Sobry A, Sihom F, et al. Prevalence of Buruli ulcer in Akonolinga health district, Cameroon: results of a cross sectional survey. *PLoS Negl Trop Dis*. 2009;3(6):e466.
20. Buruli ulcer disease. Relevé Épidémiologique Hebdomadaire Sect Hygiène Secrétariat Société Nations Wkly Epidemiol Rec Health Sect Secr Leag Nations. 2004 May 14;79(20):194–9.
21. Muelder K. Buruli ulcer in Benin. *The Lancet*. 1990 Nov;336(8723):1109–11.
22. Muelder K. Buruli ulcer in Benin. *Trop Doct*. 1988 Apr;18(2):53.
23. Marston BJ, Diallo MO, Horsburgh CR, Diomande I, Saki MZ, Kanga JM, et al. Emergence of Buruli ulcer disease in the Daloa region of Cote d'Ivoire. *Am J Trop Med Hyg*. 1995 Mar;52(3):219–24.
24. Marion E, Landier J, Boisier P, Marsollier L, Fontanet A, Le Gall P, et al. Geographic expansion of Buruli ulcer disease, Cameroon. *Emerg Infect Dis*. 2011 Mar;17(3):551–3.
25. Ravisse P. [Skin ulcer caused by *Mycobacterium ulcerans* in Cameroon. I. Clinical, epidemiological and histological study]. *Bull Société Pathol Exot Ses Fil*. 1977 Apr;70(2):109–24.
26. Veitch MG, Johnson PD, Flood PE, Leslie DE, Street AC, Hayman JA. A large localized outbreak of *Mycobacterium ulcerans* infection on a temperate southern Australian island. *Epidemiol Infect*. 1997 Dec;119(3):313–8.
27. Epidemiology of *Mycobacterium ulcerans* infection (Buruli ulcer) at Kinyara, Uganda. *Trans R Soc Trop Med Hyg*. 1971;65(6):763–75.
28. Barker DJ. The distribution of Buruli disease in Uganda. *Trans R Soc Trop Med Hyg*. 1972;66(6):867–74.
29. Sopoh GE, Johnson RC, Chauty A, Dossou AD, Aguiar J, Salmon O, et al. Buruli ulcer surveillance, Benin, 2003-2005. *Emerg Infect Dis*. 2007 Sep;13(9):1374–6.
30. Debacker M, Aguiar J, Steunou C, Zinsou C, Meyers WM, Scott JT, et al. *Mycobacterium ulcerans* disease: role of age and gender in incidence and morbidity. *Trop Med Int Health TM IH*. 2004 Dec;9(12):1297–304.
31. Hospers IC, Wiersma IC, Dijkstra PU, Stienstra Y, Etuaful S, Ampadu EO, et al. Distribution of Buruli ulcer lesions over body surface area in a large case series in Ghana: uncovering clues for mode of transmission. *Trans R Soc Trop Med Hyg*. 2005 Mar;99(3):196–201.
32. Bratschi MW, Bolz M, Minyem JC, Grize L, Wantong FG, Kerber S, et al. Geographic distribution, age pattern and sites of lesions in a cohort of Buruli ulcer patients from the Mapé Basin of Cameroon. *PLoS Negl Trop Dis*. 2013;7(6):e2252.
33. Raghunathan PL, Whitney EAS, Asamoah K, Stienstra Y, Taylor TH, Amofah GK, et al. Risk factors for Buruli ulcer disease (*Mycobacterium ulcerans* Infection): results from a case-control study in Ghana. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2005 May 15;40(10):1445–53.
34. Pouillot R, Matias G, Wondje CM, Portaels F, Valin N, Ngos F, et al. Risk factors for buruli ulcer: a case control study in Cameroon. *PLoS Negl Trop Dis*. 2007;1(3):e101.
35. Barker DJ. Epidemiology of *Mycobacterium ulcerans* infection. *Trans R Soc Trop Med Hyg*. 1973;67(1):43–50.

36. Boyd SC, Athan E, Friedman ND, Hughes A, Walton A, Callan P, et al. Epidemiology, clinical features and diagnosis of *Mycobacterium ulcerans* in an Australian population. *Med J Aust*. 2012 Mar 19;196(5):341–4.
37. Sopoh GE, Barogui YT, Johnson RC, Dossou AD, Makoutodé M, Anagonou SY, et al. Family relationship, water contact and occurrence of buruli ulcer in benin. *PLoS Negl Trop Dis*. 2010;4(7):e746.
38. Aiga H, Amano T, Cairncross S, Adomako J, Domako JA, Nanas O-K, et al. Assessing water-related risk factors for Buruli ulcer: a case-control study in Ghana. *Am J Trop Med Hyg*. 2004 Oct;71(4):387–92.
39. Quek TYJ, Athan E, Henry MJ, Pasco JA, Redden-Hoare J, Hughes A, et al. Risk factors for *Mycobacterium ulcerans* infection, southeastern Australia. *Emerg Infect Dis*. 2007 Nov;13(11):1661–6.
40. Nackers F, Johnson RC, Glynn JR, Zinsou C, Tonglet R, Portaels F. Environmental and health-related risk factors for *Mycobacterium ulcerans* disease (Buruli ulcer) in Benin. *Am J Trop Med Hyg*. 2007 Nov;77(5):834–6.
41. Debacker M, Portaels F, Aguiar J, Steunou C, Zinsou C, Meyers W, et al. Risk factors for Buruli ulcer, Benin. *Emerg Infect Dis*. 2006 Sep;12(9):1325–31.
42. Landier J, Boisier P, Fotso Piam F, Noumen-Djeunga B, Simé J, Wantong FG, et al. Adequate wound care and use of bed nets as protective factors against Buruli Ulcer: results from a case control study in Cameroon. *PLoS Negl Trop Dis*. 2011 Nov;5(11):e1392.
43. Toll A, Gallardo F, Ferran M, Gilaberte M, Iglesias M, Gimeno JL, et al. Aggressive multifocal Buruli ulcer with associated osteomyelitis in an HIV-positive patient. *Clin Exp Dermatol*. 2005 Nov;30(6):649–51.
44. Kibadi K, Colebunders R, Muyembe-Tamfum J-J, Meyers WM, Portaels F. Buruli ulcer lesions in HIV-positive patient. *Emerg Infect Dis*. 2010 Apr;16(4):738–9.
45. Johnson RC, Ifebe D, Hans-Moevi A, Kestens L, Houessou R, Guédénon A, et al. Disseminated *Mycobacterium ulcerans* disease in an HIV-positive patient: a case study. *AIDS Lond Engl*. 2002 Aug 16;16(12):1704–5.
46. Delaporte E, Alfandari S, Piette F. *Mycobacterium ulcerans* associated with infection due to the human immunodeficiency virus. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 1994 May;18(5):839.
47. Allen S. Buruli ulcer and HIV infection. *Int J Dermatol*. 1992 Oct;31(10):744–5.
48. Johnson RC, Nackers F, Glynn JR, de Biurrun Bakedano E, Zinsou C, Aguiar J, et al. Association of HIV infection and *Mycobacterium ulcerans* disease in Benin. *AIDS Lond Engl*. 2008 Apr 23;22(7):901–3.
49. Sopoh GE, Dossou AD, Brun LV, Barogui YT, Houézo JG, Affolabi D, et al. Severe multifocal form of buruli ulcer after streptomycin and rifampin treatment: comments on possible dissemination mechanisms. *Am J Trop Med Hyg*. 2010 Aug;83(2):307–13.
50. Pszolla N, Sarkar MR, Strecker W, Kern P, Kinzl L, Meyers WM, et al. Buruli ulcer: a systemic disease. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2003 Sep 15;37(6):e78–82.

51. Hofer M, Hirschel B, Kirschner P, Beghetti M, Kaelin A, Siegrist CA, et al. Brief report: disseminated osteomyelitis from *Mycobacterium ulcerans* after a snakebite. *N Engl J Med*. 1993 Apr 8;328(14):1007–9.
52. Portaels F, Johnson C, Aguiar J, Meyers W, Debacker M. Etude de 106 cas d'ulcères de Buruli avec atteintes osseuses traités à Zagnanado, Bénin. *Bull ALLF*. 2008 Jul;(23):48–50.
53. O'Brien DP, Comte E, Serafini M, Ehounou G, Antierens A, Vuagnat H, et al. The urgent need for clinical, diagnostic, and operational research for management of Buruli ulcer in Africa. *Lancet Infect Dis*. 2014 May;14(5):435–40.
54. Aujoulat I, Johnson C, Zinsou C, Guédénou A, Portaels F. Psychosocial aspects of health seeking behaviours of patients with Buruli ulcer in southern Benin. *Trop Med Int Health TM IH*. 2003 Aug;8(8):750–9.
55. Stienstra Y, van der Graaf WTA, Asamoah K, van der Werf TS. Beliefs and attitudes toward Buruli ulcer in Ghana. *Am J Trop Med Hyg*. 2002 Aug;67(2):207–13.
56. Merritt RW, Walker ED, Small PLC, Wallace JR, Johnson PDR, Benbow ME, et al. Ecology and transmission of Buruli ulcer disease: a systematic review. *PLoS Negl Trop Dis*. 2010;4(12):e911.
57. Wansbrough-Jones M, Phillips R. Buruli ulcer: emerging from obscurity. *Lancet*. 2006 Jun 3;367(9525):1849–58.
58. Jacobsen KH, Padgett JJ. Risk factors for *Mycobacterium ulcerans* infection. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. 2010 Aug;14(8):e677–81.
59. Johnson RC, Makoutodé M, Sopoh GE, Elsen P, Gbovi J, Pouteau LH, et al. Buruli ulcer distribution in Benin. *Emerg Infect Dis*. 2005 Mar;11(3):500–1.
60. Brou T, Broutin H, Elguero E, Asse H, Guegan J-F. Landscape diversity related to Buruli ulcer disease in Côte d'Ivoire. *PLoS Negl Trop Dis*. 2008;2(7):e271.
61. Williamson HR, Benbow ME, Campbell LP, Johnson CR, Sopoh G, Barogui Y, et al. Detection of *Mycobacterium ulcerans* in the environment predicts prevalence of Buruli ulcer in Benin. *PLoS Negl Trop Dis*. 2012 Jan;6(1):e1506.
62. Oluwasanmi JO, Solanke TF, Olurin EO, Itayemi SO, Alabi GO, Lucas AO. *Mycobacterium ulcerans* (Buruli) skin ulceration in Nigeria. *Am J Trop Med Hyg*. 1976 Jan;25(1):122–8.
63. Garchitorena A, Roche B, Kamgang R, Ossomba J, Babonneau J, Landier J, et al. *Mycobacterium ulcerans* ecological dynamics and its association with freshwater ecosystems and aquatic communities: results from a 12-month environmental survey in Cameroon. *PLoS Negl Trop Dis*. 2014 May;8(5):e2879.
64. Meyers WM, Shelly WM, Connor DH, Meyers EK. Human *Mycobacterium ulcerans* infections developing at sites of trauma to skin. *Am J Trop Med Hyg*. 1974 Sep;23(5):919–23.
65. Williamson HR, Mosi L, Donnell R, Aqqad M, Merritt RW, Small PLC. *Mycobacterium ulcerans* fails to infect through skin abrasions in a guinea pig infection model: implications for transmission. *PLoS Negl Trop Dis*. 2014 Apr;8(4):e2770.
66. Portaels F, Elsen P, Guimaraes-Peres A, Fonteyne PA, Meyers WM. Insects in the transmission of *Mycobacterium ulcerans* infection. *Lancet*. 1999 Mar 20;353(9157):986.



67. Portaels F, Meyers WM, Ablordey A, Castro AG, Chemlal K, de Rijk P, et al. First cultivation and characterization of *Mycobacterium ulcerans* from the environment. *PLoS Negl Trop Dis*. 2008;2(3):e178.
68. Roche B, Benbow ME, Merritt R, Kimbirauskas R, McIntosh M, Small PLC, et al. Identifying the Achilles' heel of multi-host pathogens: The concept of keystone "host" species illustrated by *Mycobacterium ulcerans* transmission. *Environ Res Lett ERL Web Site*. 2013;8(4):045009.
69. Marsollier L, Séverin T, Aubry J, Merritt RW, Saint André J-P, Legras P, et al. Aquatic snails, passive hosts of *Mycobacterium ulcerans*. *Appl Environ Microbiol*. 2004 Oct;70(10):6296–8.
70. Marsollier L, Stinear T, Aubry J, Saint André JP, Robert R, Legras P, et al. Aquatic plants stimulate the growth of and biofilm formation by *Mycobacterium ulcerans* in axenic culture and harbor these bacteria in the environment. *Appl Environ Microbiol*. 2004 Feb;70(2):1097–103.
71. Marion E, Eyangoh S, Yeramian E, Doannio J, Landier J, Aubry J, et al. Seasonal and regional dynamics of *M. ulcerans* transmission in environmental context: deciphering the role of water bugs as hosts and vectors. *PLoS Negl Trop Dis*. 2010;4(7):e731.
72. Marsollier L, André J-PS, Frigui W, Reyssset G, Milon G, Carbonnelle B, et al. Early trafficking events of *Mycobacterium ulcerans* within *Naucoris cimicoides*. *Cell Microbiol*. 2007 Feb;9(2):347–55.
73. Marsollier L, Aubry J, Coutanceau E, André J-PS, Small PL, Milon G, et al. Colonization of the salivary glands of *Naucoris cimicoides* by *Mycobacterium ulcerans* requires host plasmatocytes and a macrolide toxin, mycolactone. *Cell Microbiol*. 2005 Jul;7(7):935–43.
74. Marsollier L, Robert R, Aubry J, Saint André J-P, Kouakou H, Legras P, et al. Aquatic insects as a vector for *Mycobacterium ulcerans*. *Appl Environ Microbiol*. 2002 Sep;68(9):4623–8.
75. Marsollier L, Deniaux E, Brodin P, Marot A, Wondje CM, Saint-André J-P, et al. Protection against *Mycobacterium ulcerans* lesion development by exposure to aquatic insect saliva. *PLoS Med*. 2007 Feb;4(2):e64.
76. Johnson PDR, Azuolas J, Lavender CJ, Wishart E, Stinear TP, Hayman JA, et al. *Mycobacterium ulcerans* in mosquitoes captured during outbreak of Buruli ulcer, southeastern Australia. *Emerg Infect Dis*. 2007 Nov;13(11):1653–60.
77. Lavender CJ, Fyfe JAM, Azuolas J, Brown K, Evans RN, Ray LR, et al. Risk of Buruli ulcer and detection of *Mycobacterium ulcerans* in mosquitoes in southeastern Australia. *PLoS Negl Trop Dis*. 2011 Sep;5(9):e1305.
78. Wallace JR, Gordon MC, Hartsell L, Mosi L, Benbow ME, Merritt RW, et al. Interaction of *Mycobacterium ulcerans* with mosquito species: implications for transmission and trophic relationships. *Appl Environ Microbiol*. 2010 Sep;76(18):6215–22.
79. Fyfe JAM, Lavender CJ, Handasyde KA, Legione AR, O'Brien CR, Stinear TP, et al. A major role for mammals in the ecology of *Mycobacterium ulcerans*. *PLoS Negl Trop Dis*. 2010;4(8):e791.
80. Chauty A. Société Française et Francophone des Plaies et Cicatrisations - Ulcère de Buruli [Internet]. [cited 2011 Apr 30]. Available from: [http://www.sffpc.org/index.php?pg=connaiss\\_ulcere4](http://www.sffpc.org/index.php?pg=connaiss_ulcere4)
81. Trubiano JA, Lavender CJ, Fyfe JAM, Bittmann S, Johnson PDR. The incubation period of Buruli ulcer (*Mycobacterium ulcerans* infection). *PLoS Negl Trop Dis*. 2013;7(10):e2463.

82. WHO. Treatment of *Mycobacterium ulcerans* disease (Buruli Ulcer) - Guidance for health workers [Internet]. 2012 [cited 2014 Aug 19]. Available from: <http://www.who.int/buruli/treatment/en/>
83. Sizaïre V, Nackers F, Comte E, Portaels F. *Mycobacterium ulcerans* infection: control, diagnosis, and treatment. *Lancet Infect Dis*. 2006 May;6(5):288–96.
84. Johnson PDR, Stinear T, Small PLC, Pluschke G, Merritt RW, Portaels F, et al. Buruli ulcer (*M. ulcerans* infection): new insights, new hope for disease control. *PLoS Med*. 2005 Apr;2(4):e108.
85. Walsh DS, Portaels F, Meyers WM. Buruli ulcer (*Mycobacterium ulcerans* infection). *Trans R Soc Trop Med Hyg*. 2008 Oct;102(10):969–78.
86. Portaels F, Silva MT, Meyers WM. Buruli ulcer. *Clin Dermatol*. 2009 Jun;27(3):291–305.
87. Stienstra Y, van Roest MHG, van Wezel MJ, Wiersma IC, Hospers IC, Dijkstra PU, et al. Factors associated with functional limitations and subsequent employment or schooling in Buruli ulcer patients. *Trop Med Int Health TM IH*. 2005 Dec;10(12):1251–7.
88. Debacker M, Aguiar J, Steunou C, Zinsou C, Meyers WM, Portaels F. Buruli ulcer recurrence, Benin. *Emerg Infect Dis*. 2005 Apr;11(4):584–9.
89. Amofah G, Asamoah S, Afram-Gyening C. Effectiveness of excision of pre-ulcerative Buruli lesions in field situations in a rural district in Ghana. *Trop Doct*. 1998 Apr;28(2):81–3.
90. Barogui YT, Klis S, Bankolé HS, Sopoh GE, Mamo S, Baba-Moussa L, et al. Towards rational use of antibiotics for suspected secondary infections in Buruli ulcer patients. *PLoS Negl Trop Dis*. 2013;7(1):e2010.
91. Yeboah-Manu D, Kpeli GS, Ruf M-T, Asan-Ampah K, Quenin-Fosu K, Owusu-Mireku E, et al. Secondary bacterial infections of buruli ulcer lesions before and after chemotherapy with streptomycin and rifampicin. *PLoS Negl Trop Dis*. 2013;7(5):e2191.
92. O'Brien DP, Robson M, Friedman ND, Walton A, McDonald A, Callan P, et al. Incidence, clinical spectrum, diagnostic features, treatment and predictors of paradoxical reactions during antibiotic treatment of *Mycobacterium ulcerans* infections. *BMC Infect Dis*. 2013;13:416.
93. Trevillyan JM, Johnson PDR. Steroids control paradoxical worsening of *Mycobacterium ulcerans* infection following initiation of antibiotic therapy. *Med J Aust*. 2013 May 6;198(8):443–4.
94. O'Brien DP, Robson ME, Callan PP, McDonald AH. "Paradoxical" immune-mediated reactions to *Mycobacterium ulcerans* during antibiotic treatment: a result of treatment success, not failure. *Med J Aust*. 2009 Nov 16;191(10):564–6.
95. Nienhuis WA, Stienstra Y, Abass KM, Tuah W, Thompson WA, Awuah PC, et al. Paradoxical responses after start of antimicrobial treatment in *Mycobacterium ulcerans* infection. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2012 Feb 15;54(4):519–26.
96. Ruf M-T, Chauty A, Adeye A, Ardant M-F, Koussemou H, Johnson RC, et al. Secondary Buruli ulcer skin lesions emerging several months after completion of chemotherapy: paradoxical reaction or evidence for immune protection? *PLoS Negl Trop Dis*. 2011 Aug;5(8):e1252.
97. Friedman ND, McDonald AH, Robson ME, O'Brien DP. Corticosteroid Use for Paradoxical Reactions during Antibiotic Treatment for *Mycobacterium ulcerans*. *PLoS Negl Trop Dis*. 2012 Sep 27;6(9):e1767.

98. Schütte D, Pluschke G. Immunosuppression and treatment-associated inflammatory response in patients with *Mycobacterium ulcerans* infection (Buruli ulcer). *Expert Opin Biol Ther*. 2009 Feb;9(2):187–200.
99. Asiedu K, Etuaful S. Socioeconomic implications of Buruli ulcer in Ghana: a three-year review. *Am J Trop Med Hyg*. 1998 Dec;59(6):1015–22.
100. Chauty A, Ardant M-F, Adeye A, Euverte H, Guédénon A, Johnson C, et al. Promising clinical efficacy of streptomycin-rifampin combination for treatment of buruli ulcer (*Mycobacterium ulcerans* disease). *Antimicrob Agents Chemother*. 2007 Nov;51(11):4029–35.
101. Stienstra Y, Dijkstra PU, Guédénon A, Johnson RC, Ampadu EO, Mensah T, et al. Development of a questionnaire assessing Buruli ulcer-induced functional limitation. *Am J Trop Med Hyg*. 2004 Mar;70(3):318–22.
102. Ellen DE, Stienstra Y, Teelken MA, Dijkstra PU, van der Graaf WTA, van der Werf TS. Assessment of functional limitations caused by *Mycobacterium ulcerans* infection: towards a Buruli ulcer functional limitation score. *Trop Med Int Health TM IH*. 2003 Jan;8(1):90–6.
103. Renzaho AMN, Woods PV, Ackumey MM, Harvey SK, Kotin J. Community-based study on knowledge, attitude and practice on the mode of transmission, prevention and treatment of the Buruli ulcer in Ga West District, Ghana. *Trop Med Int Health TM IH*. 2007 Mar;12(3):445–58.
104. Mulder AA, Boerma RP, Barogui Y, Zinsou C, Johnson RC, Gbovi J, et al. Healthcare seeking behaviour for Buruli ulcer in Benin: a model to capture therapy choice of patients and healthy community members. *Trans R Soc Trop Med Hyg*. 2008 Sep;102(9):912–20.
105. Bratschi MW, Njih Tabah E, Bolz M, Stucki D, Borrell S, Gagneux S, et al. A case of cutaneous tuberculosis in a Buruli ulcer-endemic area. *PLoS Negl Trop Dis*. 2012;6(8):e1751.
106. Lupi O, Madkan V, Tying SK. Tropical dermatology: bacterial tropical diseases. *J Am Acad Dermatol*. 2006 Apr;54(4):559–78; quiz 578–80.
107. Lupi O, Bartlett BL, Haugen RN, Dy LC, Sethi A, Klaus SN, et al. Tropical dermatology: Tropical diseases caused by protozoa. *J Am Acad Dermatol*. 2009 Jun;60(6):897–925; quiz 926–8.
108. Lupi O, Tying SK, McGinnis MR. Tropical dermatology: fungal tropical diseases. *J Am Acad Dermatol*. 2005 Dec;53(6):931–51, quiz 952–4.
109. Silva MT, Portaels F, Pedrosa J. Pathogenetic mechanisms of the intracellular parasite *Mycobacterium ulcerans* leading to Buruli ulcer. *Lancet Infect Dis*. 2009 Nov;9(11):699–710.
110. Van der Werf TS, van der Graaf WT, Tappero JW, Asiedu K. *Mycobacterium ulcerans* infection. *Lancet*. 1999 Sep 18;354(9183):1013–8.
111. Revill WD, Morrow RH, Pike MC, Ateng J. A controlled trial of the treatment of *Mycobacterium ulcerans* infection with clofazimine. *Lancet*. 1973 Oct 20;2(7834):873–7.
112. Espey DK, Djomand G, Diomande I, Dosso M, Saki MZ, Kanga J-M, et al. A pilot study of treatment of Buruli ulcer with rifampin and dapsone. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. 2002 Mar;6(1):60–5.
113. Fehr H, Egger M, Senn I. Cotrimoxazol in the treatment of *Mycobacterium ulcerans* infection (Buruli ulcer) in west Africa. *Trop Doct*. 1994 Apr;24(2):61–3.

114. Etuaful S, Carbonnelle B, Grosset J, Lucas S, Horsfield C, Phillips R, et al. Efficacy of the combination rifampin-streptomycin in preventing growth of *Mycobacterium ulcerans* in early lesions of Buruli ulcer in humans. *Antimicrob Agents Chemother*. 2005 Aug;49(8):3182–6.
115. Nienhuis WA, Stienstra Y, Thompson WA, Awuah PC, Abass KM, Tuah W, et al. Antimicrobial treatment for early, limited *Mycobacterium ulcerans* infection: a randomised controlled trial. *Lancet*. 2010 Feb 20;375(9715):664–72.
116. Klis S, Stienstra Y, Phillips RO, Abass KM, Tuah W, van der Werf TS. Long term streptomycin toxicity in the treatment of Buruli Ulcer: follow-up of participants in the BURULICO drug trial. *PLoS Negl Trop Dis*. 2014 Mar;8(3):e2739.
117. Chauty A, Ardant M-F, Marsollier L, Pluschke G, Landier J, Adeye A, et al. Oral treatment for *Mycobacterium ulcerans* infection: results from a pilot study in Benin. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2011 Jan;52(1):94–6.
118. Friedman ND, Athan E, Hughes AJ, Khajehnoori M, McDonald A, Callan P, et al. *Mycobacterium ulcerans* disease: experience with primary oral medical therapy in an Australian cohort. *PLoS Negl Trop Dis*. 2013;7(7):e2315.
119. Huygen K, Adjei O, Affolabi D, Bretzel G, Demangel C, Fleischer B, et al. Buruli ulcer disease: prospects for a vaccine. *Med Microbiol Immunol (Berl)*. 2009 May;198(2):69–77.
120. BCG vaccination against *mycobacterium ulcerans* infection (Buruli ulcer). First results of a trial in Uganda. *Lancet*. 1969 Jan 18;1(7586):111–5.
121. Smith PG, Revill WD, Lukwago E, Rykushin YP. The protective effect of BCG against *Mycobacterium ulcerans* disease: a controlled trial in an endemic area of Uganda. *Trans R Soc Trop Med Hyg*. 1976;70(5-6):449–57.
122. Nackers F, Dramaix M, Johnson RC, Zinsou C, Robert A, DE Biurrun Bakedano E, et al. BCG vaccine effectiveness against Buruli ulcer: a case-control study in Benin. *Am J Trop Med Hyg*. 2006 Oct;75(4):768–74.
123. Portaels F, Aguiar J, Debacker M, Guédénou A, Steunou C, Zinsou C, et al. *Mycobacterium bovis* BCG vaccination as prophylaxis against *Mycobacterium ulcerans* osteomyelitis in Buruli ulcer disease. *Infect Immun*. 2004 Jan;72(1):62–5.
124. Grietens KP, Boock AU, Peeters H, Hausmann-Muela S, Toomer E, Ribera JM. “It is me who endures but my family that suffers”: social isolation as a consequence of the household cost burden of Buruli ulcer free of charge hospital treatment. *PLoS Negl Trop Dis*. 2008;2(10):e321.
125. Doig KD, Holt KE, Fyfe JAM, Lavender CJ, Eddyani M, Portaels F, et al. On the origin of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *BMC Genomics*. 2012;13:258.
126. Ang P, Rattana-Apiromyakij N, Goh CL. Retrospective study of *Mycobacterium marinum* skin infections. *Int J Dermatol*. 2000 May;39(5):343–7.
127. Stinear TP, Seemann T, Pidot S, Frigui W, Reyssset G, Garnier T, et al. Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res*. 2007 Feb;17(2):192–200.
128. Demangel C, Stinear TP, Cole ST. Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*. *Nat Rev Microbiol*. 2009 Jan;7(1):50–60.

129. Hong H, Demangel C, Pidot SJ, Leadlay PF, Stinear T. Mycolactones: immunosuppressive and cytotoxic polyketides produced by aquatic mycobacteria. *Nat Prod Rep*. 2008 Jun;25(3):447–54.
130. Stinear TP, Mve-Obiang A, Small PLC, Frigui W, Pryor MJ, Brosch R, et al. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A*. 2004 Feb 3;101(5):1345–9.
131. Hayman J. Out of Africa: observations on the histopathology of *Mycobacterium ulcerans* infection. *J Clin Pathol*. 1993 Jan;46(1):5–9.
132. Hong H, Coutanceau E, Leclerc M, Caleechurn L, Leadlay PF, Demangel C. Mycolactone diffuses from *Mycobacterium ulcerans*-infected tissues and targets mononuclear cells in peripheral blood and lymphoid organs. *PLoS Negl Trop Dis*. 2008;2(10):e325.
133. Sarfo FS, Le Chevalier F, Aka N, Phillips RO, Amoako Y, Boneca IG, et al. Mycolactone diffuses into the peripheral blood of Buruli ulcer patients--implications for diagnosis and disease monitoring. *PLoS Negl Trop Dis*. 2011 Jul;5(7):e1237.
134. George KM, Pascopella L, Welty DM, Small PL. A *Mycobacterium ulcerans* toxin, mycolactone, causes apoptosis in guinea pig ulcers and tissue culture cells. *Infect Immun*. 2000 Feb;68(2):877–83.
135. Guenin-Macé L, Veyron-Churlet R, Thoulouze M-I, Romet-Lemonne G, Hong H, Leadlay PF, et al. Mycolactone activation of Wiskott-Aldrich syndrome proteins underpins Buruli ulcer formation. *J Clin Invest*. 2013 Apr 1;123(4):1501–12.
136. Coutanceau E, Marsollier L, Brosch R, Perret E, Goossens P, Tanguy M, et al. Modulation of the host immune response by a transient intracellular stage of *Mycobacterium ulcerans*: the contribution of endogenous mycolactone toxin. *Cell Microbiol*. 2005 Aug;7(8):1187–96.
137. Simmonds RE, Lali FV, Smallie T, Small PLC, Foxwell BM. Mycolactone inhibits monocyte cytokine production by a posttranscriptional mechanism. *J Immunol Baltim Md 1950*. 2009 Feb 15;182(4):2194–202.
138. Coutanceau E, Decalf J, Martino A, Babon A, Winter N, Cole ST, et al. Selective suppression of dendritic cell functions by *Mycobacterium ulcerans* toxin mycolactone. *J Exp Med*. 2007 Jun 11;204(6):1395–403.
139. Adusumilli S, Mve-Obiang A, Sparer T, Meyers W, Hayman J, Small PLC. *Mycobacterium ulcerans* toxic macrolide, mycolactone modulates the host immune response and cellular location of *M. ulcerans* in vitro and in vivo. *Cell Microbiol*. 2005 Sep;7(9):1295–304.
140. Boulkroun S, Guenin-Macé L, Thoulouze M-I, Monot M, Merckx A, Langsley G, et al. Mycolactone suppresses T cell responsiveness by altering both early signaling and posttranslational events. *J Immunol Baltim Md 1950*. 2010 Feb 1;184(3):1436–44.
141. Guenin-Macé L, Carrette F, Asperti-Boursin F, Le Bon A, Caleechurn L, Di Bartolo V, et al. Mycolactone impairs T cell homing by suppressing microRNA control of L-selectin expression. *Proc Natl Acad Sci U S A*. 2011 Aug 2;108(31):12833–8.
142. Phillips R, Sarfo FS, Guenin-Macé L, Decalf J, Wansbrough-Jones M, Albert ML, et al. Immunosuppressive signature of cutaneous *Mycobacterium ulcerans* infection in the peripheral blood of patients with buruli ulcer disease. *J Infect Dis*. 2009 Dec 1;200(11):1675–84.

143. Fraga AG, Cruz A, Martins TG, Torrado E, Saraiva M, Pereira DR, et al. Mycobacterium ulcerans triggers T-cell immunity followed by local and regional but not systemic immunosuppression. *Infect Immun*. 2011 Jan;79(1):421–30.
144. Martins TG, Gama JB, Fraga AG, Saraiva M, Silva MT, Castro AG, et al. Local and regional re-establishment of cellular immunity during curative antibiotherapy of murine Mycobacterium ulcerans infection. *PloS One*. 2012;7(2):e32740.
145. Marion E, Song O-R, Christophe T, Babonneau J, Fenistein D, Eyer J, et al. Mycobacterial toxin induces analgesia in buruli ulcer by targeting the Angiotensin pathways. *Cell*. 2014 Jun 19;157(7):1565–76.
146. Casadevall A. Evolution of Intracellular Pathogens. *Annu Rev Microbiol*. 2008;62(1):19–33.
147. Torrado E, Fraga AG, Castro AG, Stragier P, Meyers WM, Portaels F, et al. Evidence for an intramacrophage growth phase of Mycobacterium ulcerans. *Infect Immun*. 2007 Feb;75(2):977–87.
148. Schütte D, Um-Boock A, Mensah-Quainoo E, Itin P, Schmid P, Pluschke G. Development of highly organized lymphoid structures in Buruli ulcer lesions after treatment with rifampicin and streptomycin. *PLoS Negl Trop Dis*. 2007;1(1):e2.
149. Torrado E, Fraga AG, Logarinho E, Martins TG, Carmona JA, Gama JB, et al. IFN-gamma-dependent activation of macrophages during experimental infections by Mycobacterium ulcerans is impaired by the toxin mycolactone. *J Immunol Baltim Md 1950*. 2010 Jan 15;184(2):947–55.
150. Stanford JL, Revill WD, Gunthorpe WJ, Grange JM. The production and preliminary investigation of Burulin, a new skin test reagent for Mycobacterium ulcerans infection. *J Hyg (Lond)*. 1975 Feb;74(1):7–16.
151. Dobos KM, Spotts EA, Marston BJ, Horsburgh CR, King CH. Serologic response to culture filtrate antigens of Mycobacterium ulcerans during Buruli ulcer disease. *Emerg Infect Dis*. 2000 Apr;6(2):158–64.
152. Pidot SJ, Porter JL, Marsollier L, Chauty A, Migot-Nabias F, Badaut C, et al. Serological evaluation of Mycobacterium ulcerans antigens identified by comparative genomics. *PLoS Negl Trop Dis*. 2010;4(11):e872.
153. Okenu DMN, Ofielu LO, Easley KA, Guarner J, Spotts Whitney EA, Raghunathan PL, et al. Immunoglobulin M antibody responses to Mycobacterium ulcerans allow discrimination between cases of active Buruli ulcer disease and matched family controls in areas where the disease is endemic. *Clin Diagn Lab Immunol*. 2004 Mar;11(2):387–91.
154. Diaz D, Döbeli H, Yeboah-Manu D, Mensah-Quainoo E, Friedlein A, Soder N, et al. Use of the immunodominant 18-kiloDalton small heat shock protein as a serological marker for exposure to Mycobacterium ulcerans. *Clin Vaccine Immunol CVI*. 2006 Dec;13(12):1314–21.
155. Casanova J-L, Abel L. Genetic dissection of immunity to mycobacteria: the human model. *Annu Rev Immunol*. 2002;20:581–620.
156. Guarner J, Bartlett J, Whitney EAS, Raghunathan PL, Stienstra Y, Asamo K, et al. Histopathologic features of Mycobacterium ulcerans infection. *Emerg Infect Dis*. 2003 Jun;9(6):651–6.

157. Röltgen K, Qi W, Ruf M-T, Mensah-Quainoo E, Pidot SJ, Seemann T, et al. Single nucleotide polymorphism typing of *Mycobacterium ulcerans* reveals focal transmission of buruli ulcer in a highly endemic region of Ghana. *PLoS Negl Trop Dis*. 2010;4(7):e751.
158. Stienstra Y, van der Werf TS, Oosterom E, Nolte IM, van der Graaf WTA, Etuafu S, et al. Susceptibility to Buruli ulcer is associated with the SLC11A1 (NRAMP1) D543N polymorphism. *Genes Immun*. 2006 Apr;7(3):185–9.
159. Torrado E, Adusumilli S, Fraga AG, Small PLC, Castro AG, Pedrosa J. Mycolactone-mediated inhibition of tumor necrosis factor production by macrophages infected with *Mycobacterium ulcerans* has implications for the control of infection. *Infect Immun*. 2007 Aug;75(8):3979–88.
160. Alcaïs A, Alter A, Antoni G, Orlova M, Nguyen VT, Singh M, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nat Genet*. 2007 Apr;39(4):517–22.
161. Mira MT, Alcaïs A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature*. 2004 Feb 12;427(6975):636–40.
162. Alcaïs A, Fieschi C, Abel L, Casanova J-L. Tuberculosis in children and adults: two distinct genetic diseases. *J Exp Med*. 2005 Dec 19;202(12):1617–21.
163. O’Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The immune response in tuberculosis. *Annu Rev Immunol*. 2013;31:475–527.
164. Al-Muhsen S, Casanova J-L. The genetic heterogeneity of mendelian susceptibility to mycobacterial diseases. *J Allergy Clin Immunol*. 2008 Dec;122(6):1043–51; quiz 1052–3.
165. Casanova J-L, Abel L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu Rev Genomics Hum Genet*. 2013;14:215–43.
166. Vincent QB, Ardant M-F, Adeye A, Goundote A, Saint-André J-P, Cottin J, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health*. 2014 Jul;2(7):e422–30.
167. Vincent QB, Ardant M-F, Marsollier L, Chauty A, Alcaïs A. HIV infection and Buruli ulcer in Africa. *Lancet Infect Dis*. 2014 Sep;14(9):796–7.
168. Pommelet V, Vincent Q, Ardant MF, Adeye A, Tanase A, Tondeur L, et al. Analysis of 81 osteomyelitis cases from a cohort of 1,257 PCR-confirmed *Mycobacterium ulcerans* infections in Benin. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2014 Jul 21;
169. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol*. 1999 Oct;28(5):964–74.
170. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Comput Stat Data Anal*. 2006 Aug;50:3464–85.
171. Wickham H. *ggplot2: elegant graphics for data analysis* [Internet]. Springer New York; 2009. Available from: <http://had.co.nz/ggplot2/book>
172. R Development Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. 2014. Available from: <http://www.R-project.org/>
173. UNAIDS Country Epidemiology - Benin [Internet]. 2012 [cited 2014 May 5]. Available from: <http://www.unaids.org/en/regionscountries/countries/benin/>

174. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2012 Revision, Highlights and Advance Tables. [Internet]. 2013. Available from: <http://esa.un.org/unpd/wpp/index.htm>
175. Cassisa V, Chauty A, Marion E, Ardant MF, Eyangoh S, Cottin J, et al. Use of fine needle aspiration for the diagnosis of *M. ulcerans* infection. *J Clin Microbiol* [Internet]. 2010 Apr 7; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20375229>
176. Boleira M, Lupi O, Lehman L, Asiedu KB, Kiszewski AE. Buruli ulcer. *An Bras Dermatol*. 2010 Jun;85(3):281–98; quiz 299–301.
177. Marriott I, Huet-Hudson YM. Sexual dimorphism in innate immune responses to infectious organisms. *Immunol Res*. 2006;34(3):177–92.
178. Iredell J, Whitby M, Blacklock Z. *Mycobacterium marinum* infection: epidemiology and presentation in Queensland 1971–1990. *Med J Aust*. 1992 Nov 2;157(9):596–8.
179. Casal M, Casal MM. Multicenter study of incidence of *Mycobacterium marinum* in humans in Spain. *Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis*. 2001 Feb;5(2):197–9.
180. Neyrolles O, Quintana-Murci L. Sexual Inequality in Tuberculosis. *PLoS Med*. 2009 Dec 22;6:e1000199.
181. Britton WJ, Lockwood DNJ. Leprosy. *Lancet*. 2004 Apr 10;363(9416):1209–19.
182. Yamamoto Y, Saito H, Setogawa T, Tomioka H. Sex differences in host resistance to *Mycobacterium marinum* infection in mice. *Infect Immun*. 1991 Nov;59(11):4089–96.
183. Peltola H, Pääkkönen M. Acute osteomyelitis in children. *N Engl J Med*. 2014 Jan 23;370(4):352–60.
184. Christinet V, Comte E, Ciaffi L, Odermatt P, Serafini M, Antierens A, et al. Impact of Human Immunodeficiency Virus on the Severity of Buruli Ulcer Disease: Results of a Retrospective Study in Cameroon. *Open Forum Infect Dis*. 2014 Mar 1;1(1):ofu021.
185. Bustamante J, Arias AA, Vogt G, Picard C, Galicia LB, Prando C, et al. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. *Nat Immunol*. 2011;12(3):213–21.
186. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 2002;30(1):97–101.
187. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*. 1987;236(4808):1567–70.
188. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010 Sep 2;467(7311):52–8.
189. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007 Nov;17(11):1665–74.
190. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl*. 2009 Jul 15;25(14):1754–60.



191. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010 Sep;20(9):1297–303.
192. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl.* 2009 Aug 15;25(16):2078–9.
193. Lai Y, Gallo RL. AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends Immunol.* 2009 Mar;30(3):131–41.
194. Semple F, Dorin JR.  $\beta$ -Defensins: Multifunctional Modulators of Infection, Inflammation and More? *J Innate Immun.* 2012;4(4):337–48.
195. Ahrens K, Schunck M, Podda G-F, Meingassner J, Stuetz A, Schröder J-M, et al. Mechanical and metabolic injury to the skin barrier leads to increased expression of murine  $\beta$ -defensin-1, -3, and -14. *J Invest Dermatol.* 2011 Feb;131(2):443–52.
196. Zhu C, Wang J, Cheng T, Li Q, Shen H, Qin H, et al. The potential role of increasing the release of mouse  $\beta$ -defensin-14 in the treatment of osteomyelitis in mice: a primary study. *PLoS One.* 2014;9(1):e86874.
197. Zhu C, Qin H, Cheng T, Tan H-L, Guo Y-Y, Shi S-F, et al. Staphylococcus aureus supernatant induces the release of mouse  $\beta$ -defensin-14 from osteoblasts via the p38 MAPK and NF- $\kappa$ B pathways. *Int J Mol Med.* 2013 Jun;31(6):1484–94.
198. Varoga D, Tohidnezhad M, Paulsen F, Wruck CJ, Brandenburg L, Mentlein R, et al. The role of human beta-defensin-2 in bone. *J Anat.* 2008 Dec;213(6):749–57.
199. Méndez-Samperio P. Role of antimicrobial peptides in host defense against mycobacterial infections. *Peptides.* 2008 Oct;29(10):1836–41.
200. Klug-Micu GM, Stenger S, Sommer A, Liu PT, Krutzik SR, Modlin RL, et al. CD40 ligand and interferon- $\gamma$  induce an antimicrobial response against Mycobacterium tuberculosis in human monocytes. *Immunology.* 2013 May;139(1):121–8.
201. Kisich KO, Heifets L, Higgins M, Diamond G. Antimycobacterial agent based on mRNA encoding human beta-defensin 2 enables primary macrophages to restrict growth of Mycobacterium tuberculosis. *Infect Immun.* 2001 Apr;69(4):2692–9.
202. Rivas-Santiago B, Sada E, Tsutsumi V, Aguilar-Leon D, Contreras JL, Hernandez-Pando R. beta-Defensin gene expression during the course of experimental tuberculosis infection. *J Infect Dis.* 2006 Sep 1;194(5):697–701.
203. Cervantes-Villagrana AR, Hernández-Pando R, Biragyn A, Castañeda-Delgado J, Bodogai M, Martínez-Fierro M, et al. Prime-boost BCG vaccination with DNA vaccines based in  $\beta$ -defensin-2 and mycobacterial antigens ESAT6 or Ag85B improve protection in a tuberculosis experimental model. *Vaccine.* 2013 Jan 11;31(4):676–84.
204. Cogen AL, Walker SL, Roberts CH, Hagge DA, Neupane KD, Khadge S, et al. Human beta-defensin 3 is up-regulated in cutaneous leprosy type 1 reactions. *PLoS Negl Trop Dis.* 2012;6(11):e1869.
205. Ortiz RH, Leon DA, Estevez HO, Martin A, Herrera JL, Romo LF, et al. Differences in virulence and immune response induced in a murine model by isolates of Mycobacterium ulcerans from different geographic areas. *Clin Exp Immunol.* 2009 Aug;157(2):271–81.

206. Yang D, Biragyn A, Hoover DM, Lubkowski J, Oppenheim JJ. Multiple roles of antimicrobial defensins, cathelicidins, and eosinophil-derived neurotoxin in host defense. *Annu Rev Immunol*. 2004;22:181–215.
207. Hollox EJ, Armour JAL, Barber JCK. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am J Hum Genet*. 2003 Sep;73(3):591–600.
208. Hardwick RJ, Machado LR, Zuccherato LW, Antolinos S, Xue Y, Shawa N, et al. A worldwide analysis of beta-defensin copy number variation suggests recent selection of a high-expressing DEFB103 gene copy in East Asia. *Hum Mutat* [Internet]. 2011 Mar 8 [cited 2011 Apr 19]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21387465>
209. Taudien S, Galgoczy P, Huse K, Reichwald K, Schilhabel M, Szafranski K, et al. Polymorphic segmental duplications at 8p23.1 challenge the determination of individual defensin gene repertoires and the assembly of a contiguous human reference sequence. *BMC Genomics*. 2004;5(1):92.
210. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011 May;12(5):363–76.
211. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet*. 2006 Sep;79(3):439–48.
212. Aldhous MC, Abu Bakar S, Prescott NJ, Palla R, Soo K, Mansfield JC, et al. Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Hum Mol Genet*. 2010 Dec 15;19(24):4930–8.
213. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol*. 2009;10(11):R130.
214. Abu Bakar S, Hollox EJ, Armour JAL. Allelic recombination between distinct genomic locations generates copy number diversity in human beta-defensins. *Proc Natl Acad Sci U S A*. 2009 Jan 20;106(3):853–8.
215. Ong C-T, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*. 2011 Apr;12(4):283–93.
216. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol*. 2011 juin;29(6):512–20.
217. Heward JA, Lindsay MA. Long non-coding RNAs in the regulation of the immune response. *Trends Immunol*. 2014 Jan 9;35(9):408–19.
218. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2012 Jan;13(1):36–46.
219. Cushwa WT, Medrano JF. Effects of blood storage time and temperature on DNA yield and quality. *BioTechniques*. 1993 Feb;14(2):204–7.
220. Richardson A, Narendran N, Guymer R, Vu H, Baird P. Blood storage at 4°C—factors involved in DNA yield and quality. *J Lab Clin Med*. 2006 Jun;147(6):290–4.

221. Halsall A, Ravetto P, Reyes Y, Thelwell N, Davidson A, Gaut R, et al. The quality of DNA extracted from liquid or dried blood is not adversely affected by storage at 4 °C for up to 24 h. *Int J Epidemiol*. 2008 Feb;37(Supplement 1):i7–10.
222. Di Pietro F, Ortenzi F, Tilio M, Concetti F, Napolioni V. Genomic DNA extraction from whole blood stored from 15- to 30-years at -20 °C by rapid phenol-chloroform protocol: A useful tool for genetic epidemiology studies. *Mol Cell Probes* [Internet]. 2010 Oct 26; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21029772>
223. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014 Jan 1;42(D1):D1001–6.
224. Pages H, Aboyoun P, Gentleman R and DebRoy S. Biostrings: String objects representing biological sequences, and matching algorithms. Version 2.32.1 [Internet]. 2014. Available from: <http://www.bioconductor.org/packages/release/bioc/html/Biostrings.html>
225. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010 juillet;11(7):499–511.
226. Consortium T 1000 GP. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012 Nov 1;491(7422):56–65.
227. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet*. 2009 juin;5(6):e1000529.
228. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012 août;44(8):955–9.
229. Browning SR, Browning BL. Haplotype phasing: existing methods and new developments. *Nat Rev Genet*. 2011 Oct;12(10):703–14.
230. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013 Jan;10(1):5–6.
231. Astle W, Balding DJ. Population Structure and Cryptic Relatedness in Genetic Association Studies. *Stat Sci*. 2009 Nov;24(4):451–71.
232. Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet*. 2005 Sep;1(3):e32.
233. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*. 2010 Jul;11(7):459–63.
234. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010 Oct;11(10):733–9.
235. Lawson DJ, Falush D. Population identification using genetic data. *Annu Rev Genomics Hum Genet*. 2012;13:337–61.
236. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945–59.
237. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association Mapping in Structured Populations. *Am J Hum Genet*. 2000 juillet;67(1):170–81.

238. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006 Dec;2(12):e190.
239. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006 Aug;38(8):904–9.
240. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, et al. Genes mirror geography within Europe. *Nature.* 2008 Nov 6;456(7218):98–101.
241. Intarapanich A, Shaw PJ, Assawamakin A, Wangkumhang P, Ngamphiw C, Chaichoompu K, et al. Iterative pruning PCA improves resolution of highly structured populations. *BMC Bioinformatics.* 2009;10:382.
242. Wright S. Coefficients of Inbreeding and Relationship. *Am Nat.* 1922 juillet;56(645):330–8.
243. The International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005 Oct 27;437(7063):1299–320.
244. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007 Oct 18;449(7164):851–61.
245. Wilkie AO. The molecular basis of genetic dominance. *J Med Genet.* 1994 Feb;31(2):89–98.
246. Rosendaal FR, Koster T, Vandenbroucke JP, Reitsma PH. High risk of thrombosis in patients homozygous for factor V Leiden (activated protein C resistance). *Blood.* 1995 Mar 15;85(6):1504–8.
247. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, et al. HLA and HIV-1: heterozygote advantage and B\*35-Cw\*04 disadvantage. *Science.* 1999;283(5408):1748–52.
248. Catano G, Agan BK, Kulkarni H, Telles V, Marconi VC, Dolan MJ, et al. Independent Effects of Genetic Variations in Mannose-Binding Lectin Influence the Course of HIV Disease: The Advantage of Heterozygosity for Coding Mutations. *J Infect Dis.* 2008 Jul 1;198(1):72–80.
249. Hraber P, Kuiken C, Yusim K. Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection. *Hepatology.* 2007 décembre;46(6):1713–21.
250. Thursz MR, Thomas HC, Greenwood BM, Hill AV. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat Genet.* 1997 Sep;17(1):11–2.
251. Hellemann D, Larsson A, Madsen HO, Bonde J, Jarløv JO, Wiis J, et al. Heterozygosity of mannose-binding lectin (MBL2) genotypes predicts advantage (heterosis) in relation to fatal outcome in intensive care patients. *Hum Mol Genet.* 2007 Dec 15;16(24):3071–80.
252. Penn DJ, Damjanovich K, Potts WK. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci U S A.* 2002 Aug 20;99(17):11260–4.
253. Sellis D, Callahan BJ, Petrov DA, Messer PW. Heterozygote advantage as a natural consequence of adaptation in diploids. *Proc Natl Acad Sci U S A.* 2011 Dec 20;108(51):20666–71.
254. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The Quality of Meta-Analyses of Genetic Association Studies: A Review With Recommendations. *Am J Epidemiol.* 2009 Dec 1;170(11):1333–43.
255. Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics.* 1997;53(4):1253–61.

256. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008 mai;9(5):356–69.
257. Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet Epidemiol*. 2007 May;31(4):358–62.
258. Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend Tests for Case-Control Studies of Genetic Markers: Power, Sample Size and Robustness. *Hum Hered*. 2002;53(3):146–52.
259. Consortium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–78.
260. Eu-ahsunthornwattana J, Miller EN, Fakiola M, Jeronimo SMB, Blackwell JM, Cordell HJ, et al. Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with Family-Based Data. *PLoS Genet*. 2014 Jul 17;10(7):e1004445.
261. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012 Jul;44(7):821–4.
262. Alcais A, Fieschi C, Abel L, Casanova JL. Tuberculosis in children and adults: two distinct genetic diseases. *J Exp Med*. 2005;202(12):1617–21.
263. Therneau T. *\_A Package for Survival Analysis in S\_*. R package version 2.37-7. [Internet]. 2014. Available from: <http://CRAN.R-project.org/package=survival>
264. Therneau T. *coxme: Mixed Effects Cox Models*. R package version 2.2-3 [Internet]. 2012. Available from: <http://CRAN.R-project.org/package=coxme>
265. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007 Jul;39(7):906–13.
266. Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet*. 2009 Jun;41(6):657–65.
267. Hunter DJ, Kraft P. Drinking from the Fire Hose — Statistical Issues in Genomewide Association Studies. *N Engl J Med*. 2007 Aug 2;357(5):436–9.
268. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinforma Oxf Engl*. 2005 Jan 15;21(2):263–5.
269. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, Bakker PIW de. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008 Dec 15;24(24):2938–9.
270. Teo Y-Y, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet*. 2010 Feb;11(2):149–60.
271. König IR, Loley C, Erdmann J, Ziegler A. How to include chromosome X in your genome-wide association study. *Genet Epidemiol*. 2014 Feb;38(2):97–103.
272. Ramanan VK, Shen L, Moore JH, Saykin AJ. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet TIG*. 2012 Jul;28(7):323–32.
273. Yaspan BL, Veatch OJ. Strategies for pathway analysis from GWAS data. *Curr Protoc Hum Genet* Editor Board Jonathan Haines Al. 2011 Oct;Chapter 1:Unit1.20.

274. Quintana-Murci L, Alcaïs A, Abel L, Casanova J-L. Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. *Nat Immunol*. 2007 Nov;8(11):1165–71.
275. Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet*. 2013 May;45(5):501–12.
276. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012 Jan 13;90(1):7–24.
277. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009 Oct 8;461(7265):747–53.
278. Jacquard A. Heritability: one word, three concepts. *Biometrics*. 1983 Jun;39(2):465–77.
279. Kempthorne O. Heritability: uses and abuses. *Genetica*. 1997;99(2-3):109–12.
280. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet*. 2013 Feb;14(2):139–49.
281. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet*. 2008 Apr;9(4):255–66.
282. Lewontin RC. Annotation: the analysis of variance and the analysis of causes. *Am J Hum Genet*. 1974 May;26(3):400–11.
283. Rose SPR. Commentary: heritability estimates--long past their sell-by date. *Int J Epidemiol*. 2006 Jun;35(3):525–7.
284. Kruuk LEB. Estimating genetic parameters in natural populations using the “animal model.” *Philos Trans R Soc Lond B Biol Sci*. 2004 Jun 29;359(1446):873–90.
285. Do CB, Hinds DA, Francke U, Eriksson N. Comparison of Family History and SNPs for Predicting Risk of Complex Disease. *PLoS Genet*. 2012 Oct 11;8(10):e1002973.
286. Jostins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet*. 2011 Oct 15;20(R2):R182–8.
287. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet*. 2009 Sep 15;18(18):3525–31.
288. Vassy JL, Meigs JB. Is genetic testing useful to predict type 2 diabetes? *Best Pract Res Clin Endocrinol Metab*. 2012 Apr;26(2):189–201.
289. Hirschhorn JN, Gajdos ZKZ. Genome-wide association studies: results from the first few years and potential implications for clinical medicine. *Annu Rev Med*. 2011;62:11–24.
290. Manolio TA. Bringing genome-wide association findings into clinical use. *Nat Rev Genet*. 2013 Aug;14(8):549–58.
291. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010 Aug 5;466(7307):714–9.
292. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013 Nov 7;93(5):779–97.

293. Rothman KJ. Causes. *Am J Epidemiol.* 1976 Dec 1;104(6):587–92.
294. Pacis A, Nédélec Y, Barreiro LB. When genetics meets epigenetics: deciphering the mechanisms controlling inter-individual variation in immune responses to infection. *Curr Opin Immunol.* 2014 Aug;29:119–26.
295. Ha N-T, Freytag S, Bickeboeller H. Coverage and efficiency in current SNP chips. *Eur J Hum Genet.* 2014 Sep;22(9):1124–30.







## ARTICLE 1

**Quentin B Vincent**, Marie-Françoise Ardant, Ambroise Adeye, Aimé Goundote, Jean-Paul Saint-André, Jane Cottin\*, Marie Kempf, Didier Agossadou, Christian Johnson, Laurent Abel, Laurent Marsollier, Annick Chauty, Alexandre Alcaïs

**Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study.**

Publié dans Lancet Global Health.

July 2014;2(7):e422-e430.



# Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study



Quentin B Vincent, Marie-Françoise Ardant, Ambroise Adeye, Aimé Goundote, Jean-Paul Saint-André, Jane Cottin\*, Marie Kempf, Didier Agossadou, Christian Johnson, Laurent Abel, Laurent Marsollier, Annick Chauty, Alexandre Alcaïs



## Summary

**Background** Buruli ulcer, caused by *Mycobacterium ulcerans*, was identified as a neglected emerging infectious disease by WHO in 1998. Although Buruli ulcer is the third most common mycobacterial disease worldwide, understanding of the disease is incomplete. We analysed a large cohort of laboratory-confirmed cases of Buruli ulcer from Pobè, Benin, to provide a comprehensive description of the clinical presentation of the disease, its variation with age and sex, and its effect on the occurrence of permanent functional sequelae.

**Methods** Between Jan 1, 2005, and Dec 31, 2011, we prospectively collected clinical and laboratory data from all patients with Buruli ulcer diagnosed at the Centre de Dépistage et de Traitement de l'Ulcère de Buruli in Pobè, Benin. We followed up patients to assess the frequency of permanent functional sequelae. All analyses were done on cases that were laboratory confirmed.

**Findings** 1227 cases of laboratory-confirmed Buruli ulcer were included in the analysis. Typically, patients with Buruli ulcer were children (median age at diagnosis 12 years) presenting with a unique (1172 [96%]) large ( $\geq 15$  cm, 444 [36%]) ulcerative (805 [66%]) lesion of the lower limb (733 [60%]). Atypical clinical presentation of Buruli ulcer included Buruli ulcer osteomyelitis with no identifiable present or past Buruli ulcer skin lesions, which was recorded in at least 14 patients. The sex ratio of Buruli ulcer widely varied with age, with male patients accounting for 57% ( $n=427$ ) of patients aged 15 years and younger, but only 33% ( $n=158$ ) of those older than 15 years (odds ratio [OR] 2.59, 95% CI 2.04–3.30). Clinical presentation of Buruli ulcer was significantly dependent on age and sex. 54 (9%) male patients had Buruli ulcer osteomyelitis, whereas only 28 (4%) of female patients did (OR 2.21, 95% CI 1.39–3.59). 1 year after treatment, 229 (22% of 1043 with follow-up information) patients presented with permanent functional sequelae. Presentation with oedema, osteomyelitis, or large ( $\geq 15$  cm in diameter), or multifocal lesions was significantly associated with occurrence of permanent functional sequelae (OR 7.64, 95% CI 5.29–11.31) and operationally defines severe Buruli ulcer.

**Interpretation** Our findings have important clinical implications for daily practice, including enhanced surveillance for early detection of osteomyelitis in boys; systematic search for *M. ulcerans* in osteomyelitis cases of non-specific aspect in areas endemic for Buruli ulcer; and specific disability prevention for patients presenting with osteomyelitis, oedema, or multifocal or large lesions. Our findings also suggest a crucial underestimation of the burden of Buruli ulcer in Africa and raise key questions about the contribution of environmental and physiopathological factors to the recorded heterogeneity of the clinical presentation of Buruli ulcer.

**Funding** Agence Nationale de la Recherche (ANR), Fondation Raoul Follereau, Fondation pour la Recherche Médicale (FRM), and Institut des Maladies Génétiques (IMAGINE).

**Copyright** © Vincent et al. Open Access article distributed under the terms of CC BY-NC-SA.

## Introduction

Buruli ulcer, caused by *Mycobacterium ulcerans*, is the third most common mycobacteriosis worldwide, after tuberculosis and leprosy.<sup>1</sup> It mostly affects rural areas of tropical countries. Although no official estimate of global incidence is available at present, west Africa is the main endemic zone, with more than 4000 new cases reported by Côte d'Ivoire, Ghana, and Benin in 2010.<sup>2</sup> Buruli ulcer is a devastating necrotising skin infection classically characterised by preulcerative lesions (nodules, plaques, oedematous infiltration), which develop into deep ulcers with undermined edges that can spread to an entire limb and disseminate to the bone. Osteomyelitis occurs in

5–10% of patients with Buruli ulcer and is usually concomitant to a skin lesion.<sup>3</sup> In some patients, most of whom are children, Buruli ulcer causes lifelong functional sequelae.

This disease was first reported in the late 19th century, but became a public health problem in the 1980s. WHO formally identified Buruli ulcer as a neglected emerging tropical disease with the launch of its global initiative against Buruli ulcer in 1998.<sup>4</sup> The increase in the number of cases was too large to be attributed exclusively to a previous lack of awareness and has been explained in several instances by local environmental events, such as changes in forest cover, flooding, the building of dams,

Lancet Glob Health 2014;  
e422–30

Published Online  
June 18, 2014  
[http://dx.doi.org/10.1016/S2214-109X\(14\)70223-2](http://dx.doi.org/10.1016/S2214-109X(14)70223-2)

See [Comment](#) page e371

\*Died on Aug 18, 2011

Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Recherche Médicale U1163 (INSERM), Paris, France (Q B Vincent MD, L Abel MD, A Alcaïs MD); Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris, France (Q B Vincent, L Abel, A Alcaïs); Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB), Pobè, Benin (M-F Ardant MD, A Adeye MD, A Goundote MD, A Chauty MD); Fondation Raoul Follereau, Paris, France (M-F Ardant, A Adeye, A Goundote, C Johnson MD, A Chauty); Laboratoire d'Anatomie Pathologique, CHU d'Angers, Angers, France (J-P Saint-André MD); Laboratoire de Bactériologie, CHU d'Angers, Angers, France (J Cottin PhD, M Kempf PhD); Programme de Lutte Contre la Lèpre et l'Ulcère de Buruli, Ministère de la Santé, Cotonou, Bénin (D Agossadou MD); St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA (L Abel, A Alcaïs); Institut National de la Recherche Médicale U892 (INSERM) et CNRS U6299, Université et CHU d'Angers, Angers, France (L Marsollier PhD); and Unité de Recherche Clinique, Paris Centre Descartes Necker Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France (A Alcaïs)

Correspondence to:  
Dr Alexandre Alcaïs, Laboratory of Human Genetics of Infectious Diseases, INSERM U1163, Imagine Institute, 75015 Paris, France  
[alexandre.alcais@inserm.fr](mailto:alexandre.alcais@inserm.fr)

or irrigation.<sup>5</sup> Several risk factors for Buruli ulcer have been repeatedly identified and include proximity to stagnant or slow-flowing bodies of water, poor wound care, and not wearing protective clothing (eg, long-sleeved shirts or shoes).<sup>6–9</sup> The exact mode of transmission is unclear, and might involve direct transmission of the bacterium from the environment to wounded skin or indirect transmission mediated by a biting vector insect.<sup>10–12</sup>

The first comprehensive epidemiological description of Buruli ulcer dates from 1971, with the seminal report on the prospective follow-up of epidemics in the Kinyara settlement, in which about 10% of 2500 initially unaffected refugees developed the disease after moving into this zone of endemic Buruli ulcer in Uganda.<sup>13</sup> This study remains unique in its almost experimental setting and it led to the first analysis of the incubation period for Buruli ulcer, the age-specific and sex-specific incidence of Buruli ulcer, and the effect of the distance between the patients' dwelling and the water source on risk of Buruli ulcer. The 1999 Ghana national case search<sup>14</sup> included the largest number of patients (>5000) studied worldwide so far and provided information about the age and sex of patients and the site of lesions. The Beninese Zagnanado study assessed 1630–2399 cases, but did not clearly specify the proportion of these cases confirmed by laboratory tests.<sup>3,15,16</sup> A short report described the distribution of age, sex, site, and type of lesions for 2598 patients attending four centres in Benin, including Pobè, in 2003–05.<sup>17</sup> Finally, a series of 750 cases focused on the site of Buruli ulcer lesions.<sup>18</sup> Other studies have provided valuable descriptions of smaller case series (100–300 patients; appendix).

Although of great interest, each of these studies had substantial limitations, such as the use of a purely descriptive approach (ie, no measurement of association), a large proportion of diagnoses being retrospective and scar-based or prospective but only clinical, or the inclusion of a large proportion of individuals for whom key data were missing. WHO insists on laboratory confirmation of

clinically suspected Buruli ulcer, with possible differential diagnoses including other tropical ulcers, skin fungal infections, or cutaneous tuberculosis.<sup>1</sup> The most sensitive and specific test is the detection of *M ulcerans* DNA by PCR on fine-needle aspiration, or biopsy or swab samples.<sup>1,19</sup> Direct smear examination after Ziehl-Neelsen staining, culture, or histopathology can also be used. In the four largest epidemiological studies of Buruli ulcer (ie, ≥750 individuals), laboratory confirmation of diagnosis was available for about 50% of cases at best.<sup>13–15,18</sup> In this study, we analyse a large cohort of laboratory-confirmed cases of Buruli ulcer from Pobè, Benin, to provide a robust and comprehensive description of the clinical presentation of Buruli ulcer, of its variation with age and sex, and its effect on the frequency of permanent functional sequelae as assessed after systematic follow-up.

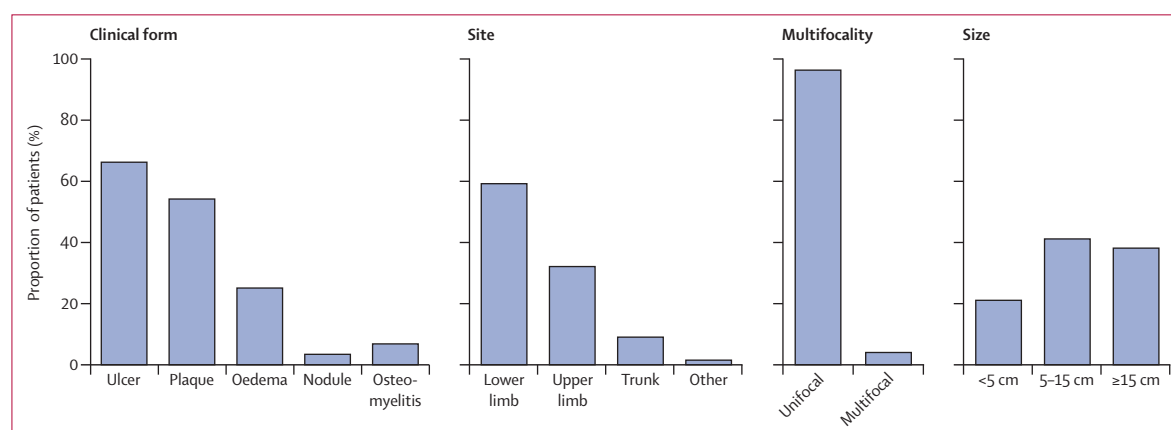
## Methods

### Participants and study design

Between Jan 1, 2005, and Dec 31, 2011, we prospectively collected clinical and laboratory data from all consecutive patients treated for Buruli ulcer at the Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB) in Pobè, Benin. We prospectively recorded age at diagnosis, sex, geographic origin of the patient, clinical form (nodule, plaque, oedema, ulcer, osteomyelitis), WHO lesion size category (maximum diameter <5 cm, 5–15 cm, ≥15 cm), and site (right or left, upper or lower limb, thorax, abdomen, head, perineum), laboratory confirmation test (culture, Ziehl-Neelsen staining or highly specific IS2404 PCR), HIV status, medical or surgical treatment, and time to healing. Notably, a patient with Buruli ulcer might have lesions at one or more sites (defining multifocality—eg, right and left arms), but also several forms at one site (eg, an ulcerated plaque).

We followed up the patients of our cohort to assess the frequency of permanent functional sequelae, ranging from the loss of ten or more degrees of joint mobility to the amputation of a whole limb. Access to the registry

See Online for appendix



**Figure 1: Clinical presentation of Buruli ulcer (Pobè, Benin, laboratory-confirmed cases, 2005–11)**

Note that a patient with Buruli ulcer might have lesions at one or more sites (defining multifocality—eg, right and left arms), but also several forms at a single site (eg, an ulcerated plaque).

was approved by the institutional review board of the CDTUB and the national Buruli ulcer control authorities.

In accordance with WHO recommendations for diagnosis of Buruli ulcer (ie, confirmation of diagnosis by at least one laboratory test; appendix), we focused our descriptive and analytical studies on laboratory-confirmed cases only. We described the clinical presentation of Buruli ulcer and analysed the effect of age at diagnosis and sex on this presentation. Ten elements of clinical presentation and outcome were assessed: male sex; localisation of a lesion on the upper body; presence of a nodule, plaque, ulcer, oedema, or osteomyelitis; largest lesion with a maximum diameter exceeding 15 cm; presence of multifocal lesions; and the frequency of permanent functional sequelae. The upper body was defined as the head, thorax, abdomen, and arms, and the lower body as the perineum and legs.

### Statistical analysis

We systematically fitted three logistic regression models, one with age, one with sex, and a third with both age and sex as explanatory variables for each of the ten elements of Buruli ulcer clinical presentation and outcome. We then screened each element of clinical presentation of Buruli ulcer for their effect on the frequency of permanent functional sequelae in both univariable and multivariable models. We operationally defined severe Buruli ulcer as presenting with at least one clinical element at higher risk for permanent functional sequelae. We further validated our definition by assessing the effect of Buruli ulcer severity on time to healing (Wilcoxon test). Age was analysed as a continuous variable and modelled by means of multiple fractional polynomials (with corresponding p values referred to as MFP-p).<sup>20</sup> For ease of interpretation, odds ratios (OR) for dichotomised age with a cutoff point at 15 years were also reported (with corresponding OR referred to as OR<sub>15</sub>). Parameter estimation and significance testing were done within the classical maximum likelihood framework. The analyses were done with R statistical analysis software (version 3.1.0, *glm* functions), together with the additional *mfp* (version 1.4.9) and *ggplot2* (version 0.9.3.1) packages.<sup>21–23</sup>

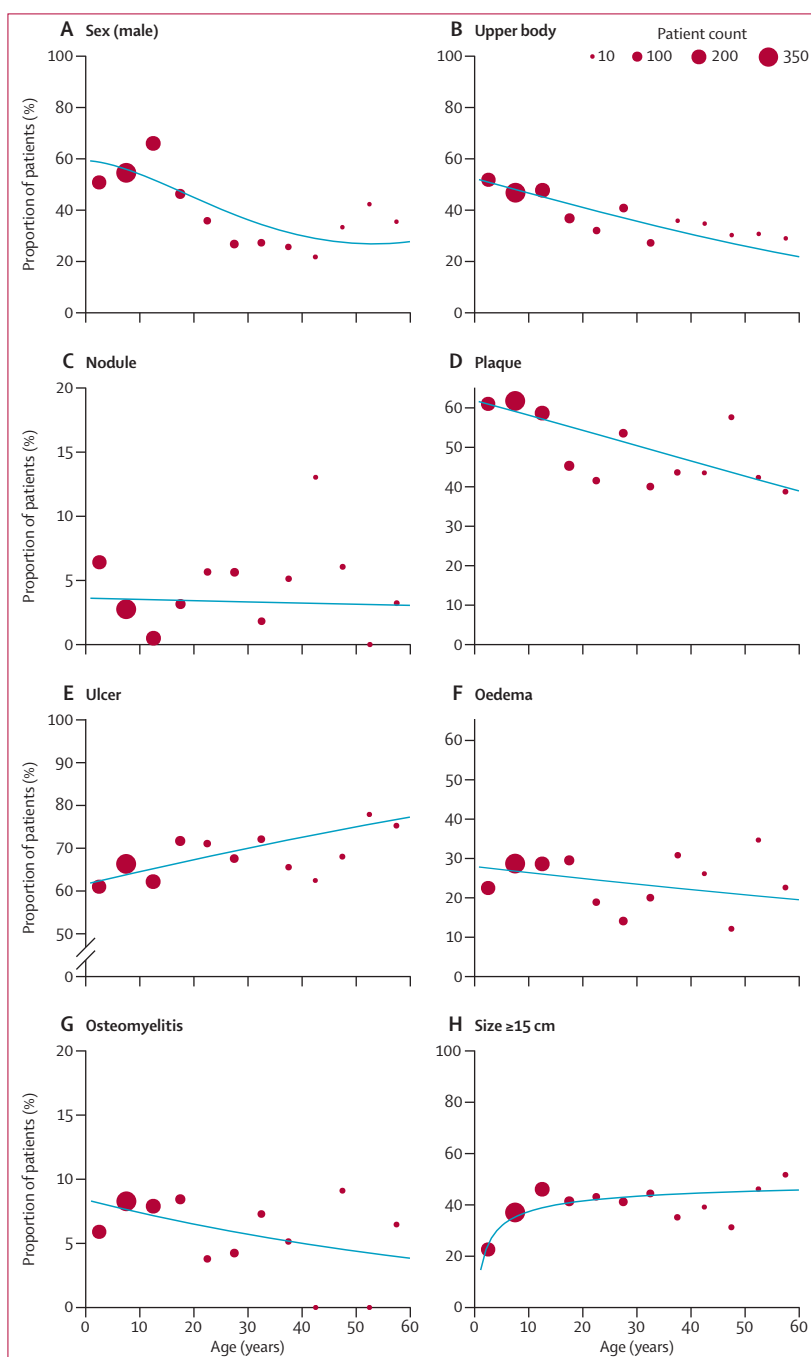
### Role of the funding source

The funders of the study had no role in study design; data collection, data analysis, data interpretation; writing of the report or in the decision to submit the paper for publication. AAL had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Results

1511 patients with Buruli ulcer were treated between Jan 1, 2005, and Dec 31, 2011, at the CDTUB, Pobè, Benin. The appendix shows the distribution of cases according to year of diagnosis, month of diagnosis, and management characteristics (antibiotic therapy, surgery, hospital admission, and time to healing). 1251 (83%) cases were

confirmed by one or more laboratory tests. PCR results were positive in 1177 (78%) cases, Ziehl-Neelsen staining was positive in 842 (56%) cases, and culture was positive in 214 (14%) cases (appendix). To ensure maximum robustness and homogeneity in our dataset, we focused



**Figure 2:** Clinical presentation of Buruli ulcer as a function of age (Pobè, Benin, laboratory-confirmed cases, 2005–11)

Red circles show the percentage of individuals with the clinical characteristic by age groups with a width of 5 years. The area of circles are scaled to the total number of individuals in the age group (ranging from about 20 to 350). The blue lines show the prediction of the logistic regression model, as fitted by the fractional polynomial method. Total counts of individuals in each regression model were 1163 for size and 1224 for all other analyses.

our analyses on 1227 laboratory-confirmed cases without HIV co-infection. We describe Buruli ulcer clinical presentation, analyse the effect of age and sex on Buruli ulcer clinical presentation, and assess the relation between clinical presentation and the occurrence of permanent

functional sequelae, incidentally leading to the proposal of an operational definition of severe Buruli ulcer.

Patients with typical Buruli ulcer were children presenting with one large ulcerative lesion of the lower limb, as detailed below (figure 1). The median age at diagnosis was 12 years (IQR 7–28 years; mean 19·3 years). Comparison with the median age of the Beninese population (18 years in 2010<sup>24</sup>) shows that Buruli ulcer is over-represented in children. 1172 (96%) patients presented with one localisation and 444 (36%) patients presented with a lesion of more than 15 cm in diameter (figure 1). 733 (60%) patients presented with a lesion on the lower limb. 805 (66%) patients had ulcers, 668 (54%) had plaques, 307 (25%) had oedema, and 42 (3%) had nodules. 82 (7%) patients had osteomyelitis (figure 1), of whom at least 14 (17%) presented with no identifiable present or past Buruli ulcer skin lesions (notably, several of these patients had fistulisation of the bone infection to the skin, as commonly recorded in osteomyelitis).

Although the overall sex ratio of the patients was balanced (640 [52%] women vs 587 [48%] men,  $p=0\cdot13$ ), a major distortion of the sex ratio was recorded as a function of age, with males being predominant in younger patients and females in older patients ( $OR_{15}$  2·59, 95% CI 2·04–3·30, MFP- $p<0\cdot0001$ , figure 2A). Specifically, male patients accounted for 427 (57%) of the patients younger than 15 years, but only 158 (33%) of those older than 15 years (table 1). This effect cannot be accounted for by the demography of the country because the Beninese population shows a balanced sex ratio both in patients younger than 15 years and in those older than 15 years.<sup>24</sup> Thus, age at diagnosis was substantially different between the male and female patients with Buruli ulcer: the median age at diagnosis was 10 years (mean 16 years) for male patients and 15 years (22·5 years) for female patients ( $p<0\cdot0001$ ).

Table 1 and figure 2 show the complete set of results regarding clinical presentation as a function of age and sex. 54 (9%) men had osteomyelitis, but only 28 (4%) women did ( $OR$  2·21, 95% CI 1·39–3·59;  $p=0\cdot0007$ ; table 1). Age was significantly associated with the frequency of clinical forms such as developing lesions on the upper body, presenting with a plaque, and presenting with an ulcer (table 1). Younger patients were prone to developing lesions on the upper body ( $OR_{15}$  2·00, 95% CI 1·58–2·55, MFP- $p<0\cdot0001$ ; figure 2) and to present with a plaque (1·90, 1·51–2·40; MFP- $p<0\cdot0001$ ; figure 2D). Conversely, older patients were more likely to present with ulcers ( $OR_{15}$  0·68, 95% CI 0·53–0·88; MFP- $p=0\cdot0004$ ; figure 2E). Additionally, we noted a number of borderline significant associations; nodules were more frequent in female patients ( $OR$  0·53, 95% CI 0·27–1·01,  $p=0\cdot05$ ; table 1) than in male patients, and osteomyelitis was more frequent in younger patients than in older patients ( $OR_{15}$  1·46, 0·91–2·41, MFP- $p=0\cdot06$ ; figure 2G).

Follow-up information was available for a median of 359 days for 1043 (85%) patients, 229 (22%) of whom

Patients, n (%)		Univariable*		Bivariable†	
		Crude OR (95% CI)	p value‡	Adjusted OR (95% CI)	p value
<b>Male sex</b>					
Age					
Continuous	..	..	<0·0001	..	..
>15 years	158 (34%)	1	..	..	..
≤15 years	427 (57%)	2·59 (2·04–3·30)	<0·0001	..	..
<b>Upper body lesion</b>					
Age					
Continuous	..	..	<0·0001	..	<0·0001
>15 years	150 (32%)	1	..	1	..
≤15 years	364 (48%)	2·00 (1·58–2·55)	<0·0001	2·13 (1·67–2·74)	<0·0001
Sex					
Female	275 (43%)	1	..	1	..
Male	239 (41%)	0·91 (0·73–1·14)	0·4241	0·78 (0·61–0·98)	0·0366
<b>Nodule</b>					
Age					
Continuous	..	..	0·7496	..	0·4886
>15 years	19 (4%)	1	..	1	..
≤15 years	23 (3%)	0·75 (0·40–1·41)	0·3639	0·85 (0·45–1·62)	0·6247
Sex					
Female	28 (4%)	1	..	1	..
Male	14 (2%)	0·53 (0·27–1·01)	0·0528	0·51 (0·26–0·99)	0·0427
<b>Plaque</b>					
Age					
Continuous	..	..	<0·0001	..	<0·0001
>15 years	211 (45%)	1	..	1	..
≤15 years	457 (61%)	1·90 (1·51–2·40)	<0·0001	2·06 (1·62–2·63)	<0·0001
Sex					
Female	361 (56%)	1	..	1	..
Male	307 (52%)	0·85 (0·68–1·06)	0·1491	0·75 (0·60–0·95)	0·0167
<b>Ulcer</b>					
Age					
Continuous	..	..	0·0004	..	<0·0001
>15 years	333 (71%)	1	..	1	..
≤15 years	469 (62%)	0·68 (0·53–0·88)	0·0024	0·64 (0·49–0·82)	0·0004
Sex					
Female	405 (63%)	1	..	1	..
Male	400 (68%)	1·24 (0·98–1·57)	0·0731	1·36 (1·07–1·74)	0·0120
<b>Oedema</b>					
Age					
Continuous	..	..	0·0445	..	0·0413
>15 years	103 (22%)	1	..	1	..
≤15 years	204 (27%)	1·33 (1·01–1·75)	0·0390	1·35 (1·02–1·78)	0·0349
Sex					
Female	160 (25%)	1	..	1	..
Male	147 (25%)	1·00 (0·77–1·30)	0·9863	0·95 (0·73–1·24)	0·7214

(Table 1 continues on next page)

presented with permanent functional sequelae, including ten amputations (<1%; five on the upper limb, five on the lower limb, ranging from one finger to a whole limb). Age and sex were not significantly associated with the frequency of permanent functional sequelae (table 1), whereas patients presenting with oedema, bone lesions, large lesions (more than 15 cm of diameter), or multifocal lesions were at significantly increased risk of permanent functional sequelae (table 2). On the basis of this finding, we propose an operational definition of a severe case of Buruli ulcer as a patient presenting with at least one of these clinical characteristics. 616 (50%) patients had severe Buruli ulcer. 192 (37%) patients with severe Buruli ulcer developed permanent functional sequelae compared with 37 (7%) patients with non-severe Buruli ulcer (OR 7.64, 95% CI 5.29–11.31,  $p<0.0001$ ; figure 3). A multivariable regression analysis showed some dependencies between the four clinical components of severity: only osteomyelitis and lesion size had independent significant effects on sequelae (osteomyelitis, OR 6.48, 95% CI 2.27–20.9,  $p=0.0004$ ; lesion size, 6.95, 4.70–10.4,  $p<0.0001$ ). In further validation of the relevance of our definition of severity, the median time to healing increased from 81 days in patients with non-severe Buruli ulcer to 107 days in patients with severe Buruli ulcer ( $p<0.0001$ ).

## Discussion

Despite its discovery more than a century ago and its rapid progression since the 1980s, Buruli ulcer remains one of the most neglected infectious diseases. With our unique collection of data recorded prospectively during 7 years at a leading treatment centre for Buruli ulcer (appendix), we have generated a comprehensive epidemiological description of Buruli ulcer in a west African setting, including systematic follow-up for permanent functional sequelae. This work provides insight into the clinical presentation and outcome of Buruli ulcer and proposes guidelines for future research efforts (panel). By contrast with other studies, this study was based on a large number of prospectively diagnosed laboratory-confirmed cases (95% of which by PCR) with a very low proportion of missing data throughout the analyses. Several characteristics of the disease consistently reported in previous studies were confirmed (appendix), a strong support to the generalisability of our findings: Buruli ulcer being mainly a paediatric disease (median age of 12 years at diagnosis); the preponderance of unifocal lesions (>90%); the predominance of lesions on the lower limbs (about 60%); ulcer as the most frequent clinical presentation (about 70%); a not unsubstantial proportion of patients presenting with osteomyelitis (>5%); and a high frequency of permanent functional sequelae (>20%). Of the many additional findings reported in this study, five are of particular importance and further discussed below: the variation of Buruli ulcer sex ratio with age; the over-representation of

		Patients, n (%)	Univariable*		Bivariable†	
			Crude OR (95% CI)	p value‡	Adjusted OR (95% CI)	p value
(Continued from previous page)						
Osteomyelitis						
Age						
Continuous	..	..		0.0560	..	0.1993
>15 years	25 (5%)	1		..	1	..
≤15 years	57 (8%)	1.46 (0.91–2.41)	0.1181		1.23 (0.76–2.05)	0.4106
Sex						
Female	28 (4%)	1	..		1	..
Male	54 (9%)	2.21 (1.39–3.59)	0.0007		2.09 (1.30–3.38)	0.0019
Lesion size ≥15 cm						
Age						
Continuous	..	..		<0.0001	..	<0.0001
>15 years	189 (42%)	1		..	1	..
≤15 years	255 (36%)	0.78 (0.61–0.99)	0.0422		0.74 (0.57–0.94)	0.0157
Sex						
Female	226 (36%)	1	..		1	..
Male	218 (40%)	1.18 (0.93–1.49)	0.1746		1.28 (1.00–1.63)	0.0466
Multifocal lesions						
Age						
Continuous	..	..		0.1292	..	0.1308
>15 years	18 (4%)	1		..	1	..
≤15 years	35 (5%)	1.23 (0.69–2.24)	0.4878		1.22 (0.68–2.26)	0.5020
Sex						
Female	27 (4%)	1	..		1	..
Male	26 (4%)	1.05 (0.60–1.82)	0.8654		0.97 (0.55–1.70)	0.9114
Severe form						
Age						
Continuous	..	..		0.0003	..	0.0002
>15 years	245 (52%)	1		..	1	..
≤15 years	371 (49%)	0.90 (0.71–1.13)	0.3537		0.85 (0.67–1.08)	0.1736
Sex						
Female	306 (48%)	1	..		1	..
Male	310 (53%)	1.22 (0.97–1.53)	0.0842		1.27 (1.01–1.59)	0.0416
Permanent sequelae						
Age						
Continuous	..	..		0.1098	..	0.2006
>15 years	79 (20%)	1		..	1	..
≤15 years	150 (23%)	1.20 (0.88–1.63)	0.2479		1.13 (0.83–1.55)	0.4425
Sex						
Female	109 (20%)	1	..		1	..
Male	120 (24%)	1.30 (0.97–1.75)	0.0783		1.26 (0.93–1.70)	0.1289
OR=odds ratio. *Univariable analysis of each clinical characteristic as a function of the predictors age and sex. †Bivariable analysis of each clinical characteristic as a function of the predictors age and sex. ‡Age was analysed both as a continuous variable modelled with multiple fractional polynomials (MFP) and as a categorical variable with a cutoff point at 15 years (figure 2 shows MFP visualisation and appendix shows MFP formulas).						
Table 1: Effect of age at diagnosis and sex on clinical presentation and outcome of Buruli ulcer (laboratory-confirmed cases, 2005–11, Pobè, Benin): univariable and bivariable analysis						

boys among patients with Buruli ulcer osteomyelitis; the existence of Buruli ulcer osteomyelitis patients with no identifiable present or past Buruli ulcer skin lesions; the strong association between clinical presentation and the



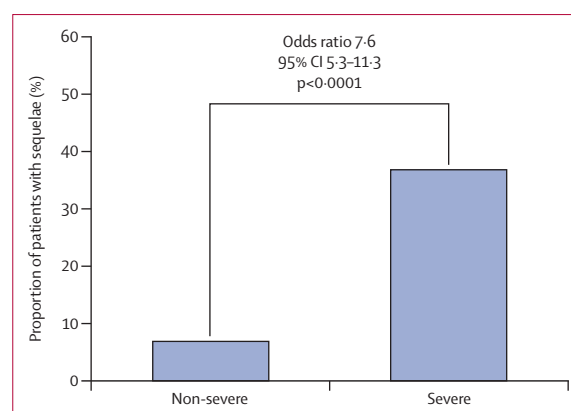
	n (%) with permanent sequelae	Univariable*		Multivariable†	
		Crude OR (95% CI)	p value	Adjusted OR (95% CI)	p value
<b>Upper body</b>					
Unaffected	124 (21%)	1	..	1	..
Affected	105 (24%)	1.22 (0.91–1.63)	0.1930	1.59 (1.10–2.30)	0.0128
<b>Nodule</b>					
Absence	227 (23%)	1	..	1	..
Presence	2 (5%)	0.20 (0.03–0.65)	0.0045	0.47 (0.07–1.73)	0.2829
<b>Plaque</b>					
Absence	143 (31%)	1	..	1	..
Presence	86 (15%)	0.40 (0.29–0.54)	<0.0001	0.64 (0.45–0.92)	0.0156
<b>Ulcer</b>					
Absence	69 (20%)	1	..	1	..
Presence	160 (23%)	1.24 (0.91–1.71)	0.1758	2.07 (1.33–3.28)	0.0012
<b>Oedema</b>					
Absence	146 (19%)	1	..	1	..
Presence	83 (31%)	1.89 (1.38–2.59)	<0.0001	1.39 (0.90–2.15)	0.1347
<b>Osteomyelitis</b>					
Absence	185 (19%)	1	..	1	..
Presence	44 (66%)	8.18 (4.87–14.1)	<0.0001	6.48 (2.27–20.9)	0.0004
<b>Size</b>					
<15 cm	47 (8%)	1	..	1	..
≥15 cm	153 (41%)	8.30 (5.82–12.0)	<0.0001	6.95 (4.70–10.4)	<0.0001
<b>Multifocal</b>					
No	210 (21%)	1	..	1	..
Yes	19 (49%)	3.59 (1.87–6.87)	0.0002	0.91 (0.37–2.18)	0.8328
<b>Severe form</b>					
No	37 (7%)	1	..	..	..
Yes	192 (37%)	7.64 (5.29–11.3)	<0.0001	..	..

OR=odds ratio. \*Univariable analysis of permanent functional sequelae as a function of each of the nine predictors.  
†Multivariable analysis of permanent functional sequelae was done as a function of all clinical characteristics (except severity).

**Table 2: Effect of clinical presentation of Buruli ulcer on permanent functional sequelae (laboratory-confirmed cases, 2005–11, Pobè, Benin)**

development of permanent functional sequelae; and the underestimation of the burden of Buruli ulcer in Africa.

Although overall balanced, the sex ratio varied widely with age. Male patients accounted for significantly more than half of the patients diagnosed younger than 15 years, but only a third of those diagnosed older than 15 years. Such sex ratio variation with age has been previously reported,<sup>7,13,14,25,26</sup> but surprisingly has always been disregarded or even denied in review papers.<sup>1,6,19,27–29</sup> Sexual dimorphism is frequently recorded in human infectious diseases.<sup>31</sup> In the context of human mycobacteriosis, rare infections with *Mycobacterium marinum* or *Mycobacterium avium intracellulare* and common infections with *Mycobacterium leprae* and *Mycobacterium tuberculosis* are more frequent in adult men than they are in adult women.<sup>31–35</sup> Explanatory hypotheses include differential exposure, differential host response, or both, according to



**Figure 3: Frequency of permanent functional sequelae in non-severe Buruli ulcer versus severe Buruli ulcer (Pobè, Benin, laboratory-confirmed cases, 2005–11)**

Severe Buruli ulcer was defined as presenting with at least one of the following clinical elements: oedema, osteomyelitis, multifocal lesions, or large lesions (≥15 cm).

sex. Experimental studies comparing the inoculation of intact male mice, female mice, and castrated male mice have shown that high testosterone concentrations impair resistance to *M. marinum*.<sup>36</sup> In our observational study in natura, infection with *M. ulcerans* followed the opposite pattern in adults, showing that despite its close relation to *M. marinum*, the toxin-secreting *M. ulcerans* has a distinctive physiopathology. This information is of particular interest and substantiates the need for specific studies that aim to understand the mechanisms underlying the age-dependent dynamics of the sex ratio in Buruli ulcer.

The risk of developing Buruli ulcer osteomyelitis was significantly greater in male patients than in female patients (OR 2.21, 95% CI 1.39–3.59), and we also recorded borderline significant association with younger age (figure 2G, table 1). No effects of age and sex on the occurrence of Buruli ulcer osteomyelitis were detected in a previous study of BCG efficacy in Buruli ulcer.<sup>37</sup> In view of our results, this difference could mean that the previous study had insufficient power to detect these effects—the power to detect an OR of 2.2 in a sample of 186 male patients and 187 female patients, assuming a type I error of 0.05, is less than 40%.<sup>37</sup> Because of its crucial clinical and physiopathological relevance, this result needs to be confirmed in independent well powered studies of Buruli ulcer, but such an excess of males has also been reported in non-Buruli ulcer osteomyelitis.<sup>38</sup> Close monitoring of boys with Buruli ulcer should be undertaken to detect osteomyelitis as early as possible. Understanding why boys are prone to Buruli ulcer osteomyelitis (eg, hormonal or host genetics components) is of major physiopathological interest and, again, should be the focus of future studies.

We report for the first time, to our knowledge, cases of Buruli ulcer osteomyelitis with no identifiable present or past Buruli ulcer skin lesions and estimated that these so-called exclusive Buruli ulcer osteomyelitis cases represent

about 20% of all Buruli ulcer osteomyelitis cases—a lower bound estimate in view of the stringent criteria that we applied. These patients presented with clinical osteomyelitis that was not specific to Buruli ulcer in appearance, with *M ulcerans* identified only through systematic laboratory investigation. Arguably, recall bias could have resulted in previous small skin lesions being overlooked in some of these cases, but very likely not in all of them. Even with this oversight, this finding would remain of fundamental clinical, public health, and physiopathological relevance. The clinical relevance is obvious: *M ulcerans* causes osteomyelitis in zones endemic with Buruli ulcer, which implies that laboratory testing for the bacterium should be done and the antibiotic regimen should be adapted. With regard to physiopathology, these exclusive cases are a proof of concept that osteomyelitis in Buruli ulcer is not necessarily the result of an uncontrolled multiplication of the bacterium in its elective tissue (skin). At least some individuals are susceptible to bone invasion by a small load of bacteria without contiguous tissue destruction. Importantly, we do not question the route of inoculation of *M ulcerans* in humans—ie, we do not imply that these patients were not bitten by a vector insect or did not have a bruise that came into contact with contaminated water. However, we suggest that some patients with Buruli ulcer can develop bone lesions without an overwhelming skin reservoir for the mycobacterium. In view of the homogeneity of *M ulcerans* strains at this local scale,<sup>39</sup> this suggestion, in turn, questions the source of the interindividual variability of the human immune response to this infectious agent. We deem this is an important point to be urgently investigated through ad-hoc studies—eg, human genetic studies. This observation also supports the possible haematogenous transport of the bacterium from the site of inoculation to the bone in some patients with Buruli ulcer.

For the first time, we systematically assessed the effect of clinical presentation of Buruli ulcer on the frequency of permanent functional sequelae. Four elements of clinical presentation were associated with a longer time to healing and an increased risk of permanent functional sequelae. We thus propose an operational definition of severe Buruli ulcer as a clinical course including oedema, osteomyelitis, multifocal, or large ( $\geq 15$  cm) lesions. The estimated OR of 7.6 for the development of permanent functional sequelae in patients with severe forms of Buruli ulcer is compelling and has straightforward clinical implications. Patients presenting with any of these four types of lesions should benefit from specific clinical care, such as enhanced monitoring, intensive physiotherapy, and timely reconstruction surgery. Our finding also raises the question as to why some patients with Buruli ulcer develop severe disease. Because the delay to diagnosis was unknown in our study and previous studies, we can only conjecture that longer delay to diagnosis would account for the severity of the clinical presentation in a proportion of patients.

However, we surmise that some patients have severe Buruli ulcer despite rapid diagnosis. Identification of such patients would be of special interest to investigate variations in the host susceptibility to Buruli ulcer.

Several findings in our study suggest the probable underestimation of the burden of Buruli ulcer in Africa. The incidental observation of patients coming from Nigeria to be treated at the Pobè CDTUB (located in Benin) shed light on this phenomenon; although the annual number of cases declared by Nigeria to the WHO is consistently less than ten (eg, seven in 2010),<sup>40</sup> about 20 Nigerian cases have been treated every year in Pobè for the past 7 years. Because Nigeria is much larger than Benin, the official figures for Nigeria are therefore likely to be a severe underestimation. Additionally, so far, the failure to diagnose Buruli ulcer with atypical clinical presentation, such as exclusive Buruli ulcer osteomyelitis, also contributes to the underestimation of the burden of Buruli ulcer, because these cases are not usually tested for the presence of *M ulcerans*.

Our results raise several questions that should be addressed by specific studies in the future, most of

#### Panel: Research in context

##### Systematic review

On March 15, 2010, we searched PubMed with the search terms “buruli OR ulcerans NOT corynebacterium” and identified about 600 publications on Buruli ulcer, including many isolated case reports. Of papers with epidemiological relevance, roughly 80 were review articles, roughly 20 were case-control studies of risk factors of Buruli ulcers, and roughly 20 were cross-sectional studies or case series of more than ten patients with Buruli ulcer. This small number of heterogeneous studies prompted us to undertake a large-scale epidemiological study in Benin to present characteristics never found together in published epidemiological studies of Buruli ulcer. Our study was representative of Buruli ulcer in west Africa, where most cases arise and the number of cases is large; all cases were prospectively diagnosed and laboratory-confirmed; follow-up for permanent functional sequelae was systematic; and the clinical course of the disease was carefully recorded resulting in sparse missing data. Taking advantage of such a setting, we were able to comprehensively describe the clinical presentation of Buruli ulcer to provide a robust and powerful analysis of its variation with age and sex and to assess its effect on the occurrence of permanent functional sequelae, leading to the operational definition of severe Buruli ulcer.

##### Interpretation

Our findings are of key public health, clinical, and physiopathological relevance. With respect to public health, several of our findings suggest a crucial underestimation of the burden of Buruli ulcer in west Africa. We describe a new form of clinical presentation for Buruli ulcer, in which patients presented with osteomyelitis with no identifiable present or past Buruli ulcer skin lesions. We propose several clinical implications of our study: enhanced surveillance for the early detection of osteomyelitis in boys; systematic search for *Mycobacterium ulcerans* in osteomyelitis cases of unspecific aspect in areas endemic with Buruli ulcer; and specific disability prevention for patients presenting with oedema, osteomyelitis, multifocal or large lesions (ie, severe Buruli ulcer). Major questions were raised with regard to the environmental or biological processes underlying the recorded heterogeneity of clinical presentation of Buruli ulcer. Further studies would be of particular interest to decipher the relative contributions of the environment, exposure, lifestyle, health-care systems, microbial strains, and host genetics to the dynamics and natural history of Buruli ulcer in natural conditions.

which are in agreement with a recent review.<sup>41</sup> Studies of the dynamics of the healing of Buruli ulcer should be undertaken, with prospective collection of information designed to disentangle several situations that we could not directly differentiate in our study—namely, recurrence (a short-term infectious clinical relapse mediated by live *M ulcerans* that was incompletely eradicated), paradoxical reaction (a non-infectious clinical relapse mediated by an immunological event), and reinfection (an independent Buruli ulcer infectious episode). In several instances, we recorded new Buruli ulcer lesions that occurred several years after healing of the initial episode (appendix). By analogy with tuberculosis, this finding suggests reinfection or perhaps long-term reactivation of the contained infection. We believe that further characterisation of these patients would be particularly relevant to better our understanding of the human immune response to *M ulcerans*, especially in the prospects of vaccine development (appendix). Finally, although we ensured the homogeneity of the dataset, the monocentric design of our study restricts the number of patients for some important aspects of the disease (eg, Buruli ulcer–HIV co-infection), calling for further research. The monocentric design of our study also raises the important question of the generalisability of our findings. We believe our study to be representative of Buruli ulcer in tropical zones, where almost all Buruli ulcer cases arise. However, large series of confirmed cases from non-tropical zones, such as Australia or Japan, will be of particular interest to decipher the relative contributions of the environment, exposure, lifestyle, health-care systems, microbial strains, and host genetics to the development and dynamics of Buruli ulcer in humans.

#### Contributors

All authors were responsible for the study concept and design, and the analysis and interpretation of data. M-FA, AAd, AG, J-PSA, DA, CJ, LM, and AC collected the data. M-FA, AAd, AG, J-PSA, JC, MK, LM, and AC did or confirmed diagnostics. QV, MFA, DA, CJ, AC, and AAI acquired the data. QV, M-FA, LA, and AAI did the statistical analysis. QV, M-FA, AC, LM, LA, and AAI drafted the report. QV, LM, AC, and AAI obtained the funding.

#### Declaration of interests

We declare no competing interests.

#### Acknowledgments

We thank staff at the Centre de Diagnostic et de Traitement de l'Ulcère de Buruli (CDTUB), Pobè, Bénin; staff of the Laboratoire de Bactériologie, CHU, Angers, France; and staff from the INSERM U1163 for helpful discussions; and acknowledge support from la Fondation Raoul Follereau. QV acknowledges support from the Fondation Imagine. LM and AAI acknowledge support from the Agence Nationale de la Recherche (ANR). AAI acknowledges support from the Fondation pour la Recherche Médicale (FRM, grant number DM120091117308). AAI, LA, and LM acknowledge support from the Institut National de la Recherche et la Recherche Médicale.

#### References

- Sizaire V, Nackers F, Comte E, Portaels F. *Mycobacterium ulcerans* infection: control, diagnosis, and treatment. *Lancet Infect Dis* 2006; **6**: 288–96.
- WHO. Buruli ulcer disease factsheet. <http://www.who.int/mediacentre/factsheets/fs199/en/index.html> (accessed April 3, 2012).
- Debacker M, Aguiar J, Steunou C, et al. *Mycobacterium ulcerans* disease (Buruli ulcer) in rural hospital, Southern Benin, 1997–2001. *Emerg Infect Dis* 2004; **10**: 1391–98.
- WHO. Buruli ulcer: progress report, 2004–2008. *Wkly Epidemiol Rec* 2008; **83**: 145–54.
- Merritt RW, Walker ED, Small PLC, et al. Ecology and transmission of Buruli ulcer disease: a systematic review. *PLoS Negl Trop Dis* 2010; **4**: e911.
- Jacobsen KH, Padgett JJ. Risk factors for *Mycobacterium ulcerans* infection. *Int J Infect Dis* 2010; **14**: e677–81.
- Pouillot R, Matias G, Wondje CM, et al. Risk factors for buruli ulcer: a case control study in Cameroon. *PLoS Negl Trop Dis* 2007; **1**: e101.
- Landier J, Boissier P, Fotso Piam F, et al. Adequate wound care and use of bed nets as protective factors against Buruli ulcer: results from a case control study in Cameroon. *PLoS Negl Trop Dis* 2011; **5**: e1392.
- Marston BJ, Diallo MO, Horsburgh CR, et al. Emergence of Buruli ulcer disease in the Daloa region of Cote d'Ivoire. *Am J Trop Med Hyg* 1995; **52**: 219–24.
- Marsollier L, Robert R, Aubry J, et al. Aquatic insects as a vector for *Mycobacterium ulcerans*. *Appl Environ Microbiol* 2002; **68**: 4623–28.
- Johnson PD, Azuolas J, Lavender CJ, et al. *Mycobacterium ulcerans* in mosquitoes captured during outbreak of Buruli ulcer, southeastern Australia. *Emerg Infect Dis* 2007; **13**: 1653–60.
- Meyers WM, Shelly WM, Connor DH, Meyers EK. Human *Mycobacterium ulcerans* infections developing at sites of trauma to skin. *Am J Trop Med Hyg* 1974; **23**: 919–23.
- No authors listed. Epidemiology of *Mycobacterium ulcerans* infection (Buruli ulcer) at Kinyara, Uganda. *Trans R Soc Trop Med Hyg* 1971; **65**: 763–75.
- Amofah G, Bonsu F, Tetteh C, et al. Buruli ulcer in Ghana: results of a national case search. *Emerging Infect Dis* 2002; **8**: 167–70.
- Debacker M, Aguiar J, Steunou C, et al. *Mycobacterium ulcerans* disease: role of age and gender in incidence and morbidity. *Trop Med Int Health* 2004; **9**: 1297–304.
- Debacker M, Portaels F, Aguiar J, et al. Risk factors for Buruli ulcer, Benin. *Emerg Infect Dis* 2006; **12**: 1325–31.
- Sopoh GE, Johnson RC, Chauty A, et al. Buruli ulcer surveillance, Benin, 2003–2005. *Emerg Infect Dis* 2007; **13**: 1374–76.
- Hospers IC, Wiersma IC, Dijkstra PU, et al. Distribution of Buruli ulcer lesions over body surface area in a large case series in Ghana: uncovering clues for mode of transmission. *Trans R Soc Trop Med Hyg* 2005; **99**: 196–201.
- Wansbrough-Jones M, Phillips R. Buruli ulcer: emerging from obscurity. *Lancet* 2006; **367**: 1849–58.
- Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999; **28**: 964–74.
- Wickham H. ggplot2: elegant graphics for data analysis. Springer New York; 2009. <http://had.co.nz/ggplot2/book> (accessed May 21, 2014).
- Team RDC. R: a language and environment for statistical computing. Vienna, Austria. 2014. <http://www.R-project.org/> (accessed May 21, 2014).
- Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Computational Statistics & Data Analysis* 2006; **50**: 3464–85.
- United Nations, Department of Economic and Social Affairs, Population Division. World population prospects: the 2012 revision, highlights and advance tables, 2013. <http://esa.un.org/unpd/wpp/index.htm> (accessed May 21, 2014).
- Raghuathan PL, Whitney EAS, Asamoah K, et al. Risk factors for Buruli ulcer disease (*Mycobacterium ulcerans* infection): results from a case-control study in Ghana. *Clin Infect Dis* 2005; **40**: 1445–53.
- Barker DJ. Epidemiology of *Mycobacterium ulcerans* infection. *Trans R Soc Trop Med Hyg* 1973; **67**: 43–50.
- Portaels F, Silva MT, Meyers WM. Buruli ulcer. *Clin Dermatol* 2009; **27**: 291–305.
- Boleira M, Lupi O, Lehman L, Asiedu KB, Kiszewski AE. Buruli ulcer. *An Bras Dermatol* 2010; **85**: 281–98.
- Silva MT, Portaels F, Pedrosa J. Pathogenetic mechanisms of the intracellular parasite *Mycobacterium ulcerans* leading to Buruli ulcer. *Lancet Infect Dis* 2009; **9**: 699–710.

- 30 Marriott I, Huet-Hudson YM. Sexual dimorphism in innate immune responses to infectious organisms. *Immunol Res* 2006; **34**: 177–92.
- 31 Iredell J, Whitby M, Blacklock Z. *Mycobacterium marinum* infection: epidemiology and presentation in Queensland 1971–1990. *Med J Aust* 1992; **157**: 596–98.
- 32 Casal M, Casal MM. Multicenter study of incidence of *Mycobacterium marinum* in humans in Spain. *Int J Tuberc Lung Dis* 2001; **5**: 197–99.
- 33 Ang P, Rattana-Apiromyakit N, Goh CL. Retrospective study of *Mycobacterium marinum* skin infections. *Int J Dermatol* 2000; **39**: 343–47.
- 34 Neyrolles O, Quintana-Murci L. Sexual inequality in tuberculosis. *PLoS Med* 2009; **6**: e1000199.
- 35 Britton WJ, Lockwood DNJ. Leprosy. *Lancet* 2004; **363**: 1209–19.
- 36 Yamamoto Y, Saito H, Setogawa T, Tomioka H. Sex differences in host resistance to *Mycobacterium marinum* infection in mice. *Infect Immun* 1991; **59**: 4089–96.
- 37 Portaels F, Aguiar J, Debacker M, et al. *Mycobacterium bovis* BCG vaccination as prophylaxis against *Mycobacterium ulcerans* osteomyelitis in Buruli ulcer disease. *Infect Immun* 2004; **72**: 62–65.
- 38 Peltola H, Pääkkönen M. Acute osteomyelitis in children. *N Engl J Med* 2014; **370**: 352–60.
- 39 Röltgen K, Qi W, Ruf M-T, Mensah-Quainoo E, et al. Single nucleotide polymorphism typing of *Mycobacterium ulcerans* reveals focal transmission of buruli ulcer in a highly endemic region of Ghana. *PLoS Negl Trop Dis* 2010; **4**: e751.
- 40 WHO. Buruli ulcer epidemiological situation. [http://apps.who.int/neglected\\_diseases/ntddata/buruli/buruli.html](http://apps.who.int/neglected_diseases/ntddata/buruli/buruli.html) (accessed May 21, 2014).
- 41 O'Brien DP, Comte E, Serafini M, Et al. The urgent need for clinical, diagnostic, and operational research for management of Buruli ulcer in Africa. *Lancet Infect Dis* 2014; **14**: 435–40.

## Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed.  
We post it as supplied by the authors.

Supplement to: Vincent QB, Ardant M-F, Adeye A, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health* 2014; published online June 18. [http://dx.doi.org/10.1016/S2214-109X\(14\)70223-2](http://dx.doi.org/10.1016/S2214-109X(14)70223-2).

# Supplementary Text, Table and Figure

**Manuscript: “Clinical epidemiology of laboratory-confirmed Buruli Ulcer in Benin: a cohort study”, Vincent et al.**

In this Supplementary Section, the interested reader will find:

- **A Result and a Discussion paragraph** on the practical aspects of BU diagnosis and management at the CDTUB, Pobe, Benin, with site-specific information on the distribution of cases according to years of diagnosis, months of diagnosis, as well as management characteristics (antibiotic therapy, surgery, hospitalization, and time to healing)
- **A Supplementary Figure** that illustrates the Diagnosis and Management paragraphs
- **A Supplementary Table** that details the multiple fractional polynomial regression models that were selected for the analysis of the impact of age on BU clinical presentation and outcome.
- **A ‘Further Reading’ box**, with additional references on specific points of the main text

## **Supplementary Text: results**

**Diagnosis and management.** In total, 1511 consecutive cases were treated between 2005 and 2011 at the CDTUB, Pobe, Benin. The annual number of cases was highest in 2006 ( $n=304$ ), subsequently decreasing to about 150 cases per year in 2009-2011, despite an increase in the number of cases coming from nearby Nigeria (2005,  $n=3$  vs. 2011,  $n=19$ ) (**Supplementary Figure A**). The distribution of cases diagnosed over the year was significantly uneven with more cases diagnosed during the dry season (November-March) than during the rainy season (April-October) ( $p=0.014$ ) (**Supplementary Figure B**). Diagnosis was confirmed by at least one laboratory test in 83% of cases ( $n=1251$ ) and by two laboratory tests in 54% of cases ( $n=812$ ). Cases that were not laboratory-tested represented 3.3% ( $n=50$ ) of the cohort. PCR results were positive in 78% of cases ( $n=1177$ ), Ziehl-Neelsen staining was positive in 56% of cases ( $n=842$ ) and culture was positive in 14% of cases ( $n=214$ ) (**Supplementary Figure C**). With regards to BU osteomyelitis, fifty percent were confirmed directly on the bone lesion, either on a surgical bone sample (43%) or on liquid from a fistula (7%). The other fifty percent were either closed skin lesions with a contiguous bone lesion, which were all confirmed by fine-needle aspiration of the contiguous skin lesion (11%), or ulcers with underlying osteomyelitis that did not require bone surgery and that were confirmed on either swabs (24%) or skin biopsies (15%). 96% ( $n=1452$ ) of patients have been treated with antibiotics (streptomycin and rifampicin in 91% of cases,  $n=1377$ ), and 56% ( $n=839$ ) of patients have required surgery (ranging from simple wound debridement to excision-grafting). Sixty-nine patients were successfully treated by an exclusive oral drug treatment associating clarithromycin and rifampicin (2007-2011). The median time to healing, from diagnosis to the final removal of dressings was 88 days; about half the patients were treated exclusively as outpatients, receiving injections at outpost dispensaries overseen by the CDTUB. Twelve

cases suffered from new BU lesions more than a year (ranging from 1.5 to 5 years) after healing of the initial episode.



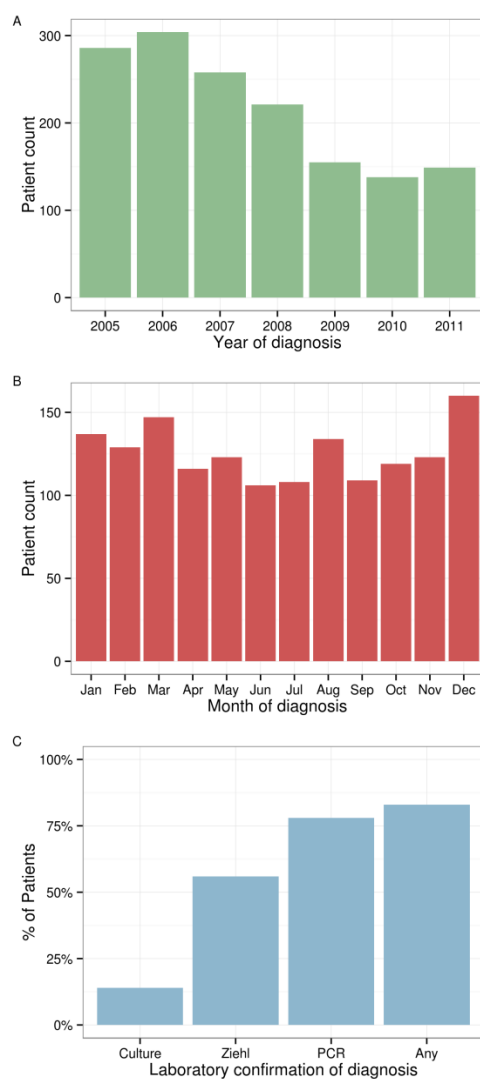
## **Supplementary Text: discussion**

Consistent with the WHO figures for Benin, the annual number of cases seen at Pobe has decreased since 2006. The reasons for this decrease are unclear, because the WHO 2005-2009 worldwide count of cases showed BU incidence to be stable at the global level (1). Some of this decrease can certainly be attributed to the exhaustive diagnosis of newly-diagnosed chronic cases following active screening campaigns in the early 2000s. This should result in a stabilization of the number of cases at the number of true incident cases. It is also possible that health promotion actions were rapidly effective, encouraging the preventive behaviors identified in several case-control studies, such as wound care, wearing protective clothes (e.g. long-sleeved clothes, shoes) and using a bed net (2–4). We cannot rule out the possibility that changes occurred in the environmental reservoir of *M. ulcerans*, or that some factors identified as having contributed to the emergence of *M. ulcerans* disappeared (5). However, we observed no clear change in the climate of the Pobe region in recent years. Another interesting hypothesis, based on the assumption that humans are themselves a reservoir for the causal bacterium, is that improvements in the medical treatment of BU may have been responsible for the decrease in the number of new cases (6). This hypothesis is compatible with the existence of an environmental step in the transmission cycle and does not imply inter-human transmission, which seems to occur only very rarely, if at all (7).

## Supplementary Text: references

1. Crompton D, World Health Organization. Working to overcome the global impact of neglected tropical diseases : first WHO report on neglected tropical diseases. Geneva Switzerland: World Health Organization; 2010.
2. Pouillot R, Matias G, Wondje CM, Portaels F, Valin N, Ngos F, et al. Risk factors for buruli ulcer: a case control study in Cameroon. *PLoS Negl Trop Dis*. 2007;1(3):e101.
3. Landier J, Boisier P, Fotso Piam F, Noumen-Djeunga B, Simé J, Wantong FG, et al. Adequate wound care and use of bed nets as protective factors against Buruli Ulcer: results from a case control study in Cameroon. *PLoS Negl Trop Dis*. November 2011;5(11):e1392.
4. Jacobsen KH, Padgett JJ. Risk factors for *Mycobacterium ulcerans* infection. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. August 2010;14(8):e677–681.
5. Merritt RW, Walker ED, Small PLC, Wallace JR, Johnson PDR, Benbow ME, et al. Ecology and transmission of Buruli ulcer disease: a systematic review. *PLoS Negl Trop Dis*. 2010;4(12):e911.
6. Sarfo FS, Lavender CJ, Fyfe JAM, Johnson PDR, Stinear TP, Phillips RO. *Mycobacterium ulcerans* DNA not detected in faecal samples from Buruli ulcer patients: results of a pilot study. *PloS One*. 2011;6(5):e19611.
7. Portaels F, Silva MT, Meyers WM. Buruli ulcer. *Clin Dermatol*. June 2009;27(3):291–305.

**Supplementary Figure. Distribution of BU cases as a function of year of diagnosis (A), month of diagnosis (B) and laboratory-confirmation status (C), Pobe, Benin (all WHO-declared cases, 2005-2011, n=1511)**



**Supplementary Table. Fractional polynomial modeling of the effect of age on BU clinical presentation and outcome, Pobe, Benin (laboratory-confirmed cases, 2005-2011)**

Clinical characteristic	Selected logistic regression model for age at diagnosis, using fractional polynomials <sup>a</sup>	Total number of individuals	Number of individuals presenting the clinical characteristic	MFP <i>p</i> -value for age <sup>b</sup>
Male	$\text{logit } P(\text{Male Sex}) = 0.38 (0.09) - 0.21 (0.04) \times \text{age}^2 + 0.10 (0.02) \text{age}^2 \times \ln(\text{age})$	1224	585	$< 10^{-4}$
Upper body lesion	$\text{logit } P(\text{Upper Body Lesion}) = 0.10 (0.09) - 0.23 (0.04) \times \text{age}$	1224	514	$< 10^{-4}$
Nodule	$\text{logit } P(\text{Nodule}) = -3.28 (0.23) - 0.03 (0.09) \times \text{age}$	1224	42	0.7496
Plaque	$\text{logit } P(\text{Plaque}) = 0.49 (0.09) - 0.16 (0.03) \times \text{age}$	1224	668	$< 10^{-4}$
Ulcer	$\text{logit } P(\text{Ulcer}) = 0.40 (0.09) + 0.13 (0.04) \times \text{age}$	1224	802	$3.7 \times 10^{-4}$
Edema	$\text{logit } P(\text{Edema}) = -0.95 (0.10) - 0.08 (0.04) \times \text{age}$	1224	307	0.0445
Osteomyelitis	$\text{logit } P(\text{Osteomyelitis}) = -2.39 (0.17) - 0.14 (0.08) \times \text{age}$	1224	82	0.0560
Size $\geq 15 \text{ cm}^c$	$\text{logit } P(\text{Size} > 15 \text{ cm}) = 0.07 (0.15) - 0.58 (0.14) \times \frac{1}{\sqrt{\text{age}}}$	1163	444	$< 10^{-4}$
Multifocality	$\text{logit } P(\text{Multifocal}) = -2.86 (0.21) - 0.14 (0.09) \times \text{age}$	1222	53	0.1292
Severity	$\text{logit } P(\text{Severe Form}) = 0.21 (0.08) - 0.17 (0.05) \times \frac{1}{\text{age}}$	1221	616	$2.7 \times 10^{-4}$
Sequelae	$\text{logit } P(\text{Sequelae}) = -1.13 (0.11) - 0.07 (0.05) \times \text{age}$	1041	229	0.1098

- a. The predicted proportion of each response is expressed as a function of the best-fitting fractional polynomial of age after application of the RA2 selection algorithm from Royston et al (1999, International Journal of Epidemiology). Coefficients (standard errors) are scaled for age expressed in decades (e.g. the prediction of the logit of the proportion of a given response at 15 years is given by replacing *age* by 1.5 in the above formulas). The predicted logit of the proportion is transformed into the predicted proportion by  $\exp(\text{logit})/(1+\exp(\text{logit}))$ , with the logit function defined as  $\ln(p/(1-p))$ .
- b. Each *p*-value corresponds to the likelihood ratio test of the given model versus the null model without age. The number of degrees of freedom depends on the form of the selected fractional polynomial for age. Here, all tests have one degree of freedom except the test of the effect of age on sex (first line, 2 degrees of freedom).

## Further Reading

**“Other studies have also provided valuable descriptions of smaller case series (100-300 patients; appendix).”**

1. Amofah GK, Sagoe-Moses C, Adjei-Acquah C, Frimpong EH. Epidemiology of Buruli ulcer in Amansie West district, Ghana. *Trans R Soc Trop Med Hyg.* December 1993;87(6):644–5.
2. Marston BJ, Diallo MO, Horsburgh CR, Diomande I, Saki MZ, Kanga JM, et al. Emergence of Buruli ulcer disease in the Daloa region of Cote d’Ivoire. *Am J Trop Med Hyg.* March 1995;52(3):219–24.
3. Raghunathan PL, Whitney EAS, Asamoah K, Stienstra Y, Taylor TH, Amofah GK, et al. Risk factors for Buruli ulcer disease (*Mycobacterium ulcerans* Infection): results from a case-control study in Ghana. *Clin Infect Dis Off Publ Infect Dis Soc Am.* May 15, 2005;40(10):1445–53.
4. Aiga H, Amano T, Cairncross S, Adomako J, Domako JA, Nanas O-K, et al. Assessing water-related risk factors for Buruli ulcer: a case-control study in Ghana. *Am J Trop Med Hyg.* October 2004;71(4):387–92.
5. Porten K, Sailor K, Comte E, Njikap A, Sobry A, Sihom F, et al. Prevalence of Buruli ulcer in Akonolinga health district, Cameroon: results of a cross sectional survey. *PLoS Negl Trop Dis.* 2009;3(6):e466.
6. Pouillot R, Matias G, Wondje CM, Portaels F, Valin N, Ngos F, et al. Risk factors for buruli ulcer: a case control study in Cameroon. *PLoS Negl Trop Dis.* 2007;1(3):e101.
7. Quek TYJ, Athan E, Henry MJ, Pasco JA, Redden-Hoare J, Hughes A, et al. Risk factors for *Mycobacterium ulcerans* infection, southeastern Australia. *Emerg Infect Dis.* November 2007;13(11):1661–6.

**“In accordance with WHO recommendations for diagnosis of Buruli ulcer (i.e. confirmation of diagnosis by at least one laboratory test; appendix)”**

WHO. Treatment of *Mycobacterium ulcerans* disease (Buruli ulcer): guidance for health workers. 2012. ISBN9789241503402.

[http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402_eng.pdf)

**“With our unique collection of data recorded prospectively during 7 years at a leading treatment centre for Buruli ulcer (appendix)”**

1. Sopoh GE, Johnson RC, Chauty A, Dossou AD, Aguiar J, Salmon O, et al. Buruli ulcer surveillance, Benin, 2003-2005. *Emerg Infect Dis*. September 2007;13(9):1374–6.
2. Chauty A, Ardant M-F, Marsollier L, Pluschke G, Landier J, Adeye A, et al. Oral treatment for *Mycobacterium ulcerans* infection: results from a pilot study in Benin. *Clin Infect Dis Off Publ Infect Dis Soc Am*. January 2011;52(1):94–6.
3. Chauty A, Ardant M-F, Adeye A, Euverte H, Guédénon A, Johnson C, et al. Promising clinical efficacy of streptomycin-rifampin combination for treatment of buruli ulcer (*Mycobacterium ulcerans* disease). *Antimicrob Agents Chemother*. November 2007;51(11):4029–35.
4. Marsollier L, Deniaux E, Brodin P, Marot A, Wondje CM, Saint-André J-P, et al. Protection against *Mycobacterium ulcerans* lesion development by exposure to aquatic insect saliva. *PLoS Med*. February 2007;4(2):e64.
5. Ruf M-T, Chauty A, Adeye A, Ardant M-F, Koussemou H, Johnson RC, et al. Secondary Buruli ulcer skin lesions emerging several months after completion of chemotherapy: paradoxical reaction or evidence for immune protection? *PLoS Negl Trop Dis*. August 2011;5(8):e1252.
6. Pidot SJ, Porter JL, Marsollier L, Chauty A, Migot-Nabias F, Badaut C, et al. Serological evaluation of *Mycobacterium ulcerans* antigens identified by comparative genomics. *PLoS Negl Trop Dis*. 2010;4(11):e872.
7. Cassisa V, Chauty A, Marion E, Ardant MF, Eyangoh S, Cottin J, et al. Use of fine needle aspiration for the diagnosis of *M. ulcerans* infection. *J Clin Microbiol* [Internet]. April 7, 2010; Retrieved from: <http://www.ncbi.nlm.nih.gov/pubmed/20375229>

**“Several characteristics of the disease consistently reported in previous studies were confirmed (appendix), a strong support to the generalisability of our findings: BU being mainly a paediatric disease (median age of 12 years at diagnosis); the preponderance of unifocal lesions (>90%) (1–5) ; the predominance of lesions on the lower limbs (>60%) (2,6,7); ulcer as the most frequent clinical presentation (~70%) (2,5); a significant proportion of patients presenting with osteomyelitis (>5%) (2,4); and a high frequency of permanent functional sequelae (>20%) (7,8).”**

1. Porten K, Sailor K, Comte E, Njikap A, Sobry A, Sihom F, et al. Prevalence of Buruli ulcer in Akonolinga health district, Cameroon: results of a cross sectional survey. *PLoS Negl Trop Dis*. 2009;3(6):e466.

2. Sopoh GE, Johnson RC, Chauty A, Dossou AD, Aguiar J, Salmon O, et al. Buruli ulcer surveillance, Benin, 2003-2005. *Emerg Infect Dis*. September 2007;13(9):1374–6.
3. Epidemiology of *Mycobacterium ulcerans* infection (Buruli ulcer) at Kinyara, Uganda. *Trans R Soc Trop Med Hyg*. 1971;65(6):763–75.
4. Debacker M, Aguiar J, Steunou C, Zinsou C, Meyers WM, Scott JT, et al. *Mycobacterium ulcerans* disease: role of age and gender in incidence and morbidity. *Trop Med Int Health TM IH*. December 2004;9(12):1297–304.
5. Hospers IC, Wiersma IC, Dijkstra PU, Stienstra Y, Etuaful S, Ampadu EO, et al. Distribution of Buruli ulcer lesions over body surface area in a large case series in Ghana: uncovering clues for mode of transmission. *Trans R Soc Trop Med Hyg*. March 2005;99(3):196–201.
6. Amofah G, Bonsu F, Tetteh C, Okrah J, Asamoah K, Asiedu K, et al. Buruli ulcer in Ghana: results of a national case search. *Emerg Infect Dis*. February 2002;8(2):167–70.
7. Marston BJ, Diallo MO, Horsburgh CR, Diomande I, Saki MZ, Kanga JM, et al. Emergence of Buruli ulcer disease in the Daloa region of Cote d’Ivoire. *Am J Trop Med Hyg*. March 1995;52(3):219–24.
8. Pouillot R, Matias G, Wondje CM, Portaels F, Valin N, Ngos F, et al. Risk factors for buruli ulcer: a case control study in Cameroon. *PLoS Negl Trop Dis*. 2007;1(3):e101.

**“We believe that further characterisation of these patients would be particularly relevant to better our understanding of the human immune response to *M. ulcerans*, especially in the prospects of vaccine development (appendix).”**

1. Huygen K, Adjei O, Affolabi D, Bretzel G, Demangel C, Fleischer B, et al. Buruli ulcer disease: prospects for a vaccine. *Med Microbiol Immunol (Berl)*. May 2009;198(2):69–77.

# Good data collection is essential for a better understanding of Buruli ulcer



Buruli ulcer is a necrotising skin and soft tissue disease caused by the environmental bacterium *Mycobacterium ulcerans*. The organism belongs to the same family as those that cause leprosy and tuberculosis, but uniquely, *M ulcerans* produces the toxin mycolactone, which causes tissue damage. Clinical manifestations include nodules, oedema, and plaques, but the most typical presentation is a large ulcer, usually on the limbs. Buruli ulcer is rarely fatal, but victims are often left deformed and with permanent disability.

Under-reporting of the disease is a major problem in several countries. Although advances in treatment have been made with WHO's recommendation<sup>1</sup> of the use of combined antibiotic treatment with rifampicin and streptomycin for 8 weeks for all types of lesion, laboratory confirmation of clinical diagnosis is still a limiting factor. The development of a PCR test that targets IS2404 (an insertion sequence found on the *M ulcerans* genome) allows accurate diagnosis, but is available only in reference laboratories that are far from areas endemic for Buruli ulcer.<sup>2</sup> Microscopy is available in local health facilities, but is less sensitive than PCR testing. To ensure standardised recording and reporting of cases, WHO developed the Buruli ulcer clinical and treatment 01 form<sup>3</sup> to collect clinical and laboratory data for all patients with suspected or confirmed Buruli ulcer.

The study by Quentin Vincent and colleagues<sup>4</sup> in *The Lancet Global Health* makes good use of data collected with the WHO Buruli ulcer forms. Vincent and colleagues exploited and analysed data for 1357 laboratory-confirmed cases of Buruli ulcer during 10 years at the treatment centre in Pobè, Benin, to provide qualitative and quantitative information about clinical features, diagnosis, treatment, and follow-up. Unlike other published studies that have included clinically diagnosed cases only, this analysis was done on laboratory-confirmed cases of Buruli ulcer, which lends significance to the study. In this study, oedema was most frequent in children, and in the most severe form, led to sequelae in 30% of cases. Plaques and nodules were also more frequent in children and all these clinical presentations decreased significantly with age. Another interesting finding is that osteomyelitis caused by

*M ulcerans* presented without any skin lesion in 20% of cases. Clinicians should therefore pay more attention to the diagnosis of osteomyelitis in young people in endemic areas. They also need to be aware of the special challenges in the diagnosis and management of Buruli ulcer according to lesion type.

Some characteristics of the disease reported in other studies were also confirmed by Vincent and colleagues: patients with Buruli ulcer were young, with a median age of 12 years, lesions were mainly localised in the limbs, and 90% of lesions were unifocal. The disseminated non-ulcerative form of *M ulcerans* infection as oedema is well known and has been reported in severe forms in children.

This study highlights the importance of using the WHO form to ensure standard collection of patient data within and across endemic countries. National programmes should encourage health workers to correctly fill out the forms and encourage supervisory activities to ensure their completeness and accuracy. This prospective series of patients is an excellent example of what can be achieved when health workers are encouraged to correctly fill out the forms. The information collected can complement other epidemiological information about patients' habits, location of housing, activities, and other health data related to the environment. The analyses of these data could further enhance understanding of risk factors and help to better develop research hypothesis on the transmission.

A key question that needs to be addressed is why children are more vulnerable to the development of non-ulcerative forms of Buruli ulcer disease such as oedema and osteomyelitis. A point-of-care test that is effective, efficient, and easy to implement on the field is urgently needed. Such a test is also essential to exclude other common lesions that could mimic Buruli ulcer and cause false-positive clinical diagnosis. A less invasive sampling method for the oedema and osteomyelitis forms of Buruli ulcer is also needed. Converse and colleagues<sup>5</sup> reported a successful innovation to detect mycolactone in tissue; future development of this technique could be used as a diagnostic test.

Despite substantial knowledge about its clinical manifestations in some endemic countries, con-

*Lancet Glob Health* 2014

Published Online  
June 18, 2014  
[http://dx.doi.org/10.1016/S2214-109X\(14\)70257-8](http://dx.doi.org/10.1016/S2214-109X(14)70257-8)

See Online/Articles  
[http://dx.doi.org/10.1016/S2214-109X\(14\)70223-2](http://dx.doi.org/10.1016/S2214-109X(14)70223-2)



firmation of Buruli ulcer diagnosis is necessary for correct management and reporting of cases. Therefore, health staff should be trained to fill out the WHO form, and physicians should be educated about Buruli ulcer. Early diagnosis and treatment, including measures to prevent disability, are essential for a good outcome.

Sara Eyangoh

Mycobacteriology Service, Centre Pasteur of Cameroon,  
Yaounde, Cameroon  
eyangoh@pasteur-yaounde.org

I declare no competing interests.

Copyright © Eyangoh. Open Access articles distributed under the terms of  
CC BY-NC-SA.

- 1 WHO. Treatment of *Mycobacterium ulcerans* disease (Buruli ulcer): guidance for health workers. World Health Organization. 2013. [http://www.who.int/iris/bitstream/10665/77771/http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402\\_eng.pdf](http://www.who.int/iris/bitstream/10665/77771/http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402_eng.pdf) (accessed May 25, 2013).
- 2 WHO. Laboratory diagnosis of Buruli ulcer: a manual for health care providers. World Health Organization. 2014. [http://www.who.int/iris/bitstream/10665/111738/http://apps.who.int/iris/bitstream/10665/111738/1/9789241505703\\_eng.pdf?ua=1](http://www.who.int/iris/bitstream/10665/111738/http://apps.who.int/iris/bitstream/10665/111738/1/9789241505703_eng.pdf?ua=1) (accessed April 16, 2014).
- 3 WHO. Buruli ulcer clinical and treatment form—new case. BU 01.N. World Health Organization. [http://www.who.int/buruli/control/ENG\\_BU\\_01\\_N.pdf?ua](http://www.who.int/buruli/control/ENG_BU_01_N.pdf?ua) (accessed May 10, 2014).
- 4 Vincent QB, Ardant M-F, Adeye A, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Global Health* 2014; published online June 18. [http://dx.doi.org/10.1016/S2214-109X\(14\)70223-2](http://dx.doi.org/10.1016/S2214-109X(14)70223-2)
- 5 Converse PJ, Xing Y, Kim KH, et al. Accelerated detection of mycolactone production and response to antibiotic treatment in a mouse model of *Mycobacterium ulcerans* disease. *PLoS Negl Trop Dis* 2014; **8**: e2618.

## ARTICLE 1

### SELECTION DES QUESTIONS DES RELECTEURS SCIENTIFIQUES (REVIEWERS)

**Quentin B Vincent**, Marie-Françoise Ardant, Ambroise Adeye, Aimé Goundote, Jean-Paul Saint-André, Jane Cottin\*, Marie Kempf, Didier Agossadou, Christian Johnson, Laurent Abel, Laurent Marsollier, Annick Chauty, Alexandre Alcaïs

**Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study.**

#### ***Themes of selected questions:***

- ***What is the difference between a relapse and a second episode?***
- ***Is BU really emerging? Awareness bias vs environmental changes?***
- ***Could you give odds-ratios for Fractional Polynomials?***
- ***Why choose 15 or 20 years old as a cutoff point for the analysis of age?***
- ***Is this really a cohort?***
- ***Are the conclusions on age and sex valid without normalization to general population?***



*The authors mention that the recurrence and/or relapse rates are very low in this centre - 'Since 2003, relapses have been recorded in 1\*8% (n=33) of patients. A second BU episode sometime after the resolution of an initial episode was observed in exceptional cases (n<5)' - and suggest this is because at this site they are proactive in their treatment policies? ('However, relapses were particularly rare at the Pobe center (<2%), possibly due to the early systematic use of antibiotic regimens at this pioneering center (28,29)'? But how usual are relapses in general? And what is the difference between a relapse and a second episode - does a new episode have to be at a different location in the body?*

We thank the reviewer for enticing us to clarify the difference between what we called a relapse and a second episode in the context of BU. A debate is ongoing in the BU community to differentiate several clinical situations where “new” BU lesions appear, reappear or extend in a healing or apparently healed BU patient. This debate currently focuses on differentiating recurrence (a short-term infectious clinical relapse mediated by live *M. ulcerans* that was incompletely eradicated; this is what we, and others, called relapses in the initial manuscript) and paradoxical reaction (a non-infectious clinical relapse mediated by an immunological event), while the notion of long-term infectious episodes has not been addressed at all (see WHO reference text below). The confusion between these situations is exemplified by the wide range of reported recurrence rates (from 6 to 47%, Sopoh et al., Emerging Infectious Diseases, 2007; Debacker et al., Emerging Infectious Diseases, 2005; Amofah et al., Tropical Doctor, 1998).

In the 2012 WHO document “Treatment of *Mycobacterium ulcerans* disease (Buruli ulcer): guidance for health workers” (ISBN9789241503402, [http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402_eng.pdf)), the definition section actually considers *new cases*, *recurrent cases* and *paradoxical reactions*, while not even considering subsequent independent infectious episodes. We cite:

“**A new case** is defined as a person presenting with a Buruli ulcer lesion who has not previously received antibiotic treatment for Buruli ulcer. **A recurrent case** is defined as a patient who has previously received antibiotics for Buruli ulcer, who presents with a lesion at another site or lesions at the same site within 1 year of the end of the last antibiotic treatment. **Paradoxical reactions** are a recently recognized phenomenon, and some cases previously classified as recurrent lesions may have been due to paradoxical reactions. Such reactions occur during or long after antibiotic treatment, with new inflammatory disease (presenting as a nodule/swelling, plaque or oedema) leading to extension of the existing ulcer or a new lesion on a different part of the body, usually with pus formation and pain. These are sometimes seen on parts of the body where there was no evidence of disease before antibiotic treatment, perhaps as a result of subclinical infection.

Cultures of tissue or pus are usually sterile, although acid-fast bacilli can still be seen and polymerase chain reaction (PCR) for *M. ulcerans* IS2404 remains positive. Histopathological examination of the lesion demonstrates an intense immunological reaction within and around the lesion.”

In our manuscript, we initially wanted to highlight a disregarded yet important clinical situation, where a patient suffers from **a new clinical infectious episode of BU** long after the resolution of a first episode. By analogy with tuberculosis, this could correspond to long-term reactivation or to reinfection by an independent strain. This is what we initially called a “second episode”. Such an observation is of major physiopathological relevance, as it implies that in at least some patients, a first episode of BU does not establish a stable protective immune response; i.e. BU does not protect against BU. This in turn has practical implications on the research and development of any BU vaccine.

However, in this context of unclear definitions across studies, we agree with the reviewer that our data on this point might not be directly generalizable and that more work is required to disentangle the three situations described above, e.g. we do not have systematic histopathology for our cases. **As a consequence, we (i) modified our initial statement to a non-mechanistic neutral one: “Twelve cases suffered from new BU lesions more than a year (from 1.5 to 5 years) after healing of the initial episode”, (ii) moved this information to supplementary material, (iii) added discussion on this point in the last paragraph of the discussion to encourage future research (Discussion Paragraph #7).**

*'The increase in the number of cases was, indeed, too massive to be attributed exclusively to an awareness bias and has been linked to environmental changes, such as changes in forest cover, flooding, the building of dams and irrigation (5)' - useful if the authors could at least estimate / speculate what magnitude of increase might be down to these environmental or methodological aspects?*

We thank the reviewer for her/his comment on this sentence of our introduction. To our knowledge, there is no available estimate of the fraction of the increase that could be explained by an obvious environmental change or by growing disease awareness. We further discuss this interesting point below.

Epidemiologists working on emerging infectious diseases are all confronted to the same difficult question: does the increase in the number of cases rather reflect a true increase in the annual number of infected individuals or a better detection of these cases (e.g. through better screening and diagnosis awareness)?

A first strategy to tackle this tough question is to consider reports of appearance and disappearance of foci of infection in places with no known change in the screening policy. Observation of new foci of infection in such a context positively answers the question: “Has BU truly emerged at least somewhere?”. Sometimes, an environmental factor could be associated to this binary emergence event (e.g. cases on a Nigerian university campus after the creation of an artificial pond, Oluwasanmi et al., American Journal of Tropical Medicine and Hygiene, 1976). In some cases, suppression of the environmental perturbation led to the extinction of the focus of infection (e.g. an Australian focus of infection around a golf course, which appeared after modification of the source of its water system, and disappeared when re-establishing the previous water source, Veitch et al, Epidemiology and Infection, 1997). This answers the question: “Has true BU emergence been convincingly associated with an environmental event at least somewhere?”. These reports provide the proof-of-principle that (i) BU foci emerged independently of the screening policy, (ii) environmental factors impact the dynamics of the disease.

Generalizing the previous examples to the interpretation of countrywide or worldwide figures of incidence is far from straightforward and by no means a specific issue of BU. In Benin for example, the first PubMed article mentioning the disease dates back to 1988 and 28 cases were reported in 1988-1989. WHO figures on the progression of the number of cases from 1989 (n=15) to 2007 (n=1200) show a ~80-fold increase in the number of cases in this country. While not impossible, we doubt that a thousand BU patients were left undiagnosed in 1989 and rather believe the increase in number of cases reflects a true emergence. Reviewer 1 proposes that we speculate on

the magnitude of the awareness bias versus true emergence: as a personal opinion, we would suggest a 5-fold increase might be attributable to better awareness. However, this is a highly speculative estimate that we are happy to share with the reviewer, but obviously much too frail to be part of the introduction of our manuscript.

As a final note, the WHO classified BU as an emerging infectious disease in 1998 (Yamoussoukro declaration) and launched the Global Initiative against BU, testifying of its belief in a true emergence of the disease.

**To clarify our view we modified the text as follows: “The increase in the number of cases was, indeed, on a too large scale to be attributed exclusively to an awareness bias and has been explained on several instances by local environmental events, such as changes in forest cover, flooding, the building of dams and irrigation (5)”**

*In the Tables of estimates, the authors report age only with a P-value. Couldn't we see the relevant parameter estimates from the fractional polynomials as well?*

As discussed in the original paper by Royston et al on fractional polynomials, the essence of fractional polynomials is to enable non-linear modelling of the effect of a continuous variable, age in the present situation. This approach has been shown to be extremely powerful and prevents the loss of information resulting from the classical dichotomization of a continuous variable (Royston et al, International Journal of Epidemiology, 1999). However, it makes user-friendly reporting of odds-ratios less straightforward. This is why we decided to also display an OR for dichotomized age.

In the context of fractional polynomials, if the selected model is linear (untransformed age as the predictor), it is easy to report an odds-ratio by, say, increases of age by ten years. If the model is non-linear, (e.g. squared age), the OR for an increase of ten years of age is not constant with age, i.e. the OR comparing the odds at 50 vs 40 years of age is not the same as the OR comparing the odds at 20 vs 10 years of age.

As exemplified in the original paper by Royston, it is possible to report a tabulation of the odds-ratios at 10, 20, ..., 60 years of age compared to a reference age. We felt this would make the table unreadable given the number of fractional polynomial models that we present. Instead, we believe that Figure 3 (now Figure 2) better displays this non-linear variation of odds with age (blue curve), while reporting dichotomized age simplified the quantification of its effect with a single OR.

**We provided full information in the Supplemental Table, with the form of the selected fractional polynomial, coefficient and standard deviations estimates, as well as sample size, to make tabulations of the fractional polynomial based ORs possible to the interested reader.**



*What is the reason of choosing the cutoff point at 20 years? Other studies, like the one in Benin by Debacker et al. which you cited estimates that people at risk are <15 and >59 years old? This information touches principally on children, since more than 50% cases are observed amongst individuals less than 15 years, knowing that they often represent a greater part of the population.*

Choosing a cutoff point in epidemiological analyses of continuous variables is always a difficult issue, with pros and cons for different possible cutoff values. This is precisely why our primary analysis of the impact of age on clinical presentation and outcome is the fractional polynomial analysis, which models age as a continuous variable instead of dichotomizing it. Interestingly, it also allows non-linear modeling the effect of age on the variable of interest, as illustrated by the blue curves in revised Figure 2 (previously Figure 3). This property of fractional polynomials however makes user-friendly reporting of odds-ratios less straightforward, which is why we decided to also display an OR for dichotomized age.

We initially chose the cutoff point at 20 years for two reasons: (i) it was a chosen cutoff in several previous studies (e.g. Amofah et al., Transactions Of The Royal Society Of Tropical Medicine And Hygiene, 1993; Amofah et al., Emerging Infectious Diseases, 2002; Ragunathan et al., Clinical Infectious Diseases, 2005), (ii) it best reproduced the statistically significant/ non-significant status of the effect of age on variables of interest from the fractional polynomial approach, which we thought would be less confusing to readers.

However, we agree that a 15 years of age cutoff should be more relevant as regards physiopathology, by better separating prepuberty and postpuberty individuals. **We therefore followed the reviewer's suggestion and changed the cut-off to 15 years.**

*"Cohort" - is the use of the term "cohort" adequate /correct in this context? Authors analyzed data well registered in a Buruli ulcer treatment center, the follow-up data registered were not always the same over the years.*

There has been confusion in the use of the terms “cohort” and “case series” in epidemiology, as nicely summarized in Dekkers et al, Annals of Internal Medicine, 2012. According to this systematic synthesis, which relied on the critical examination of 27 epidemiology textbooks (!), our study is a cohort study of BU cases who were followed-up for their functional disability status. All patients had at least one follow-up consultation about a year after healing. Therefore, the absolute risk of permanent sequelae can be estimated (and its risk factors derived), which is the distinctive characteristic for a design to be termed “cohort”. Furthermore, the follow-up data have now been homogenized in our sample: following the suggestion made by reviewers 1 and 2, we restricted the analysis to the 2005-2011 period and completed the dataset with the one-year follow-up of the patients diagnosed in 2011 (see below).

**We clarified the study design throughout the manuscript (title, abstract, methods). In the methods section we specified: “Patients of our cohort were followed-up to assess the occurrence of permanent functional sequelae”**

*The 3rd discussion paragraph on sex and age- it is difficult to assess since there is no information on the distribution of sex and age in general population*

We agree with the reviewer that this external information was missing for the valid interpretation of the median age of onset and the sex ratio of BU cases. However, this information is not required when assessing the impact of age and sex on BU-related phenotypes (see below).

**We injected external information on the distribution of age and sex in the general Beninese population in the revised manuscript.** This additional data from the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (World Population Prospects: The 2012 Revision, <http://esa.un.org/unpd/wpp/index.htm>) reinforced our initial statements:

- BU is a pediatric disease, i.e. BU cases are enriched with children as compared to the general population (median age in the Beninese population 18, vs 12 in our cohort of BU cases)
- There is an excess of males in young BU patients and of females in older BU patients (proportion of males in the general Beninese population under the age of 15: 50.3%, proportion of males in the general Beninese population over the age of 15: 49.3%)

We believe the rest of the age and sex analyses in our manuscript to be valid without external information. The analyses of the impact of age and sex on the clinical forms and outcomes of the disease (lesion type and localization, severity, sequelae) answer the following type of questions: are the clinical forms of the disease uniformly distributed among BU cases? Are bone lesions more often found in young than old patients? We do not need external data to answer these questions since here the “general” population consists of the population of BU cases only. These questions are of both physiopathological and clinical relevance and we think they qualify for consideration in the discussion.

**To avoid confusion with regards to the need of external data to interpret our results, we restructured the corresponding paragraphs of the result and discussion section into paragraphs where external data is required for interpretation (Discussion Paragraph #1 and #2) and paragraphs where internal data are sufficient (Discussion Paragraph #3).**

## ARTICLE 2

**Quentin B Vincent, Marie-Françoise Ardant, Laurent Marsollier,**

*Annick Chauty, Alexandre Alcaïs*

*on behalf of the Franco-Beninese Buruli Research Group*

**Defining and targeting high-risk populations in Buruli ulcer—Authors' reply.**

Publié dans Lancet Global Health.

November 2014;2(11):e630.





## Defining and targeting high-risk populations in Buruli ulcer

In their article in *The Lancet Global Health*, Quentin Vincent and colleagues (July issue)<sup>1</sup> analysed 1227 laboratory-confirmed cases of Buruli ulcer and provided clinically relevant results such as an operational classification of lesions. They reported unbalanced age and sex groups, as previously noted,<sup>2</sup> and an over-representation of Buruli ulcer cases in children. Quantification of this over-representation is challenging—proper comparisons between age groups need estimation of age-specific incidence rates, which are rarely reported.<sup>2,3</sup> Furthermore, most studies of risk factors rely on age-matched and sex-matched case-control designs, precluding comparative analysis by age and sex.<sup>4</sup>

Here we report incidence rates for men and women by 5 year age-groups for 814 laboratory-confirmed cases in the Nyong River region in Cameroon (figure). Estimates were obtained using Buruli ulcer surveillance data from the

Centre Pasteur du Cameroun (National Reference Laboratory for Mycobacteria, Yaoundé, Cameroon) from 2002 to 2012, and detailed 2010 population census data. Incidence rate was twice as high for children aged 5–14 years (165 per 100 000 person-years) as for adults older than 15 years (87 per 100 000 person-years). Incidence rate was stable among adults, suggesting lifelong exposure and possibly increasing susceptibility to disease development among older individuals. The overall women–men incidence rate ratio was 1·01 (95% CI 0·88–1·15). Incidence rate ratio was 0·81 (0·67–0·98) for children aged 0–14 years, 1·55 (1·22–1·98) for adults aged 15–49 years, and 0·85 (0·57–1·28) for adults older than 50 years.

Vincent and colleagues<sup>1</sup> defined age-specific and sex-specific subgroups relevant for clinical considerations.<sup>1</sup> We define three subgroups—boys younger than 15 years, women aged 15–50 years, and elderly people—which are likely to have different exposures, behaviours, immunity, and treatment-seeking attitudes. We advocate that this risk categorisation is relevant for

epidemiological studies and public health programmes such as prevention or case-detection campaigns.

We declare no competing interests.

Copyright © Landier et al. Open Access article distributed under the terms of CC BY-NC-ND.

\*Jordi Landier, Arnaud Fontanet, Gaëtan Texier

jordi.landier@gmail.com

Institut Pasteur, Unité de Recherche et d'expertise en Épidémiologie des Maladies Emergentes, Paris, France (JL, AF); Service d'Épidémiologie et de Santé Publique, Centre Pasteur du Cameroun, Réseau International des Instituts Pasteur, Yaoundé, Cameroon (JL, GT); Chaire Santé et Développement, Conservatoire National des Arts et Métiers, Paris, France (AF); Aix-Marseille Université, UMR912 SESSTIM (INSERM-IRD-AMU), Marseille 13005, France (GT)

- 1 Vincent QB, Ardant MF, Adeye A, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health* 2014; **2**: e422–30.
- 2 Debacker M, Aguiar J, Steunou C, et al. *Mycobacterium ulcerans* disease: role of age and gender in incidence and morbidity. *Trop Med Int Health* 2004; **9**: 1297–304.
- 3 Revill WLD, Barker DJP. Seasonal distribution of mycobacterial skin ulcers. *Brit J Prev Soc Med* 1972; **26**: 23–27.
- 4 Jacobsen KH, Padgett JJ. Risk factors for *Mycobacterium ulcerans* infection. *Int J Infect Dis* 2010; **14**: e677–81.

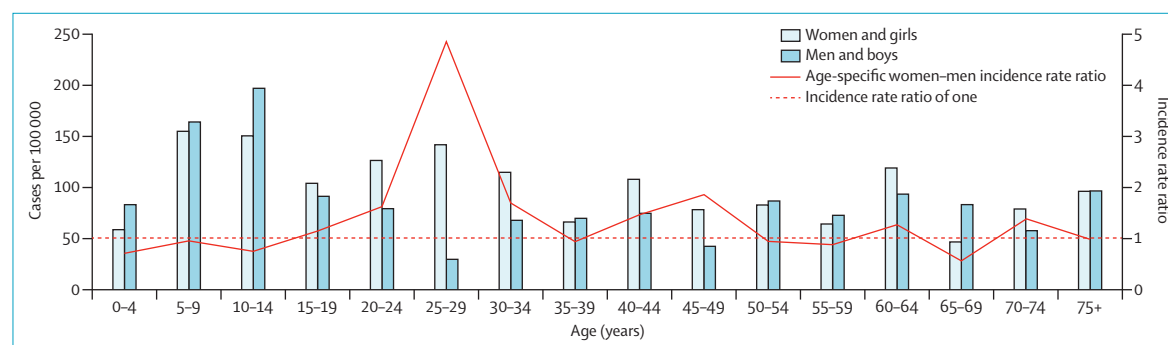


Figure: Incidence rate and women–men incidence rate ratio of Buruli ulcer in the Nyong River valley (cases per 100 000 person-years), 2002–12, Cameroon



## Defining and targeting high-risk populations in Buruli ulcer

### Authors' reply

We thank Jordi Landier and colleagues for their comments about our recent Article in *The Lancet Global Health*.<sup>1</sup> In their work, Landier and coworkers generalise some of our observations on Buruli ulcer in Benin to those for Cameroon, the country that has the fifth highest prevalence of Buruli ulcer worldwide. Briefly, they make use of age and sex distribution from the Cameroon national census to show that patients aged 5–14 years were twice as likely to be affected by Buruli ulcer as older individuals; and that boys were over-represented in individuals younger than 15 years, women were over-represented in patients aged 15–50 years, and that men and women were equally represented in patients older than 50 years. They advocate the use of national census references to produce incidence rates and incidence rate ratios (IRRs), which they believe to be the proper way to draw valid conclusions.

We agree that this method is an appropriate way to assess whether Buruli ulcer is over-represented in children, but this issue was not central to our study. We also felt that the introduction of external data, the quality of which we cannot assess and which might not be relevant to areas where Buruli ulcer is endemic—ie, remote rural areas of tropical countries—could lead to bias. Nevertheless, our Article reports that the median age of the population in Benin in 2010 is significantly higher than the median age of our cohort,<sup>1</sup> leading to the same conclusion as the IRRs. Landier and colleagues also define

elderly people as a high incidence group. We would be cautious about this conclusion because their report and others do not provide the number of patients in each age group to assess the uncertainty of their measure.<sup>2</sup> As an example, consideration of the low number of patients older than 60 years in our study (one of the largest worldwide) enticed us not to draw conclusions in that age category. With respect to the variation of the sex ratio with age, we are mindful that the national census correction is correct but negligible, because the general population sex ratio does not differ from 1 in Benin, whatever the age group considered.<sup>3</sup> Therefore, Landier and colleagues reach the same overall conclusions as us, an extremely valuable confirmation.

Landier and colleagues note, as we did, that unbalanced age-gender distribution among patients with Buruli ulcer had previously been reported. However, the reference they quote is misleading, because it does not identify differential incidence by sex before the age of 60 years.<sup>2</sup> The research group stated in a later 2009 review that “there are no sex differences in disease instance among children and adults”,<sup>4</sup> as did another independent review cited by Landier and colleagues.<sup>5</sup> This is regrettable because this important issue of the variation of the sex ratio with age, although repeatedly reported in research articles on Buruli ulcer, has been disregarded or even denied by most authors of review papers<sup>4,5</sup> (see discussion in our Article<sup>1</sup> for details). This denial has far too long hampered the research on the causes of this age-dependent variation of the sex-ratio in Buruli ulcer, understanding of which will be of great physiopathological, clinical,

and, as pointed out by Landier and colleagues, public health relevance.

We declare no competing interests. We thank staff at the Centre de Dépistage et de Traitement de l'Ulcère de Buruli, Pobè, Benin; staff of the Laboratoire de Bactériologie, CHU, Angers, France; and staff from the Institut National de la Santé et de la Recherche Médicale (INSERM) U1163 for helpful discussions; and acknowledge support from la Fondation Raoul Follereau. QBV acknowledges support from the Fondation Imagine. LM and AA acknowledge support from the Agence Nationale de la Recherche. AA acknowledges support from the Fondation pour la Recherche Médicale (grant number DMI20091117308). AA and LM acknowledge support from INSERM.

Copyright © Vincent et al. Open Access article distributed under the terms of CC BY-NC-SA.

Quentin B Vincent,  
Marie-Françoise Ardant,  
Laurent Marsollier, Annick Chauty,  
\*Alexandre Alcaïs, for the  
Franco-Beninese Buruli Research  
Group (members listed in appendix)  
alexandre.alcais@inserm.fr

Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Recherche Médicale U1163, Paris, France (QBV, AA); Institut Imagine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France (QBV, AA); Centre de Dépistage et de Traitement de l'Ulcère de Buruli, Pobè, Benin (M-FA, AC); Fondation Raoul Follereau, Paris, France (M-FA, AC); Institut National de la Recherche Médicale U892 et CNRS U6299, Université et CHU d'Angers, Angers, France (LM); St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA (AA), and Unité de Recherche Clinique, Paris Centre Descartes Necker Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France (AA)

- 1 Vincent QB, Ardant MF, Adeye A, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health* 2014; **2**: e422–30.
- 2 Debacker M, Aguiar J, Steunou C, et al. *Mycobacterium ulcerans* disease: role of age and gender in incidence and morbidity. *Trop Med Int Health* 2004; **9**: 1297–304.
- 3 UN Department of Economic and Social Affairs, Population division. World population prospects: The 2012 revision; highlights and advance tables. 2013. <http://esa.un.org/unpd/wpp/index.htm> (accessed Sept 30, 2014).
- 4 Portaels F, Silva MT, Meyers WM. Buruli ulcer. *Clin Dermatol* 2009; **27**: 291–305.
- 5 Jacobsen KH, Padgett JJ. Risk factors for *Mycobacterium ulcerans* infection. *Int J Infect Dis* 2010; **14**: e677–81.



See Online for appendix





## ARTICLE 3

**Quentin B Vincent**, Marie-Françoise Ardant, Laurent Marsollier,

Annick Chauty, Alexandre Alcaïs

*on behalf of the Franco-Beninese Buruli Research Group*

**HIV infection and Buruli ulcer in Africa.**

Publié dans Lancet Infectious Diseases.

September 2014;14(9):796–7.





models of innovation must be explored.

Médecins Sans Frontières, in consultation with other stakeholders, has developed the "3P Project" (Push, Pull, Pool),<sup>5</sup> an incentive framework for tuberculosis drug regimen development. The 3P Project aims to deliver affordable, effective new regimens for tuberculosis more efficiently, through an open, collaborative approach to drug development, and through novel approaches to finance and coordination of research and development, including push funding (ie, through grants), pull funding (ie, through milestone prizes), and pooling of intellectual property to ensure open collaborative research and fair licensing for high-quality low-cost production of the final products.

We agree with Zumla and colleagues that the specialty cannot be complacent, and that to speed up regimen development and re-engage the private sector in tuberculosis research and development novel approaches are necessary.

We declare no competing interests.

**Manica Balasegaram,**  
\*Grania Brigden  
grania.brigden@geneva.msf.org

Access Campaign, Médecins Sans Frontières,  
78 Rue de Lausanne, Geneva, Switzerland

- 1 Zumla A I, Gillespie S H, Hoelscher M, et al. New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. *Lancet Infect Dis* 2014; **14**: 327–40.
- 2 Brigden G, Nyang'wa BT, du Cros P, et al. Principles for designing future regimens for multidrug-resistant tuberculosis. *Bull World Health Org* 2014; **92**: 68–74.
- 3 Stop TB Partnership, Working Group on New TB Drugs. Drug Pipeline. 2014. <http://www.newtbdrugs.org/pipeline.php> (accessed March 28, 2014).
- 4 Treatment Action Group. 2013 report on tuberculosis research funding trends, 2005–2012. 2013. <http://www.treatmentactiongroup.org/tbrd2013> (accessed March 28, 2014).
- 5 Médecins Sans Frontières Access Campaign. PUSH, PULL, POOL: accelerating innovation and access to medicines for tuberculosis. 2014. <http://www.msfaccess.org/push-pull-pool> (accessed March 28, 2014).

## HIV infection and Buruli ulcer in Africa

In a recent Personal View on Buruli ulcer, Daniel O'Brien and colleagues<sup>1</sup> highlighted the urgent need for research on HIV–Buruli ulcer co-infection. Whether HIV infection is a risk factor for Buruli ulcer is unknown, because the few studies published show contrasting results. In Ghana, HIV was not associated with Buruli ulcer (six of 116 Buruli ulcer cases with HIV versus one of 116 controls,  $p=0.89$ ), whereas a significant association was found in Benin (11 of 426 versus two of 613,  $p=0.003$ ).<sup>2,3</sup> This discrepancy could result from HIV being differentially associated with subgroups of Buruli ulcer, such as severe Buruli ulcer. Sporadic case reports described patients with HIV with both severe and non-severe evolution, for example.<sup>4,5</sup> However, as emphasised by O'Brien and colleagues, no published epidemiological data exist on the association between HIV and severity of Buruli ulcer.

We did a large cohort study in a highly endemic region of Benin, and addressed the question of the effect of HIV on the severity of Buruli ulcer. Clinical and laboratory data were prospectively obtained from all consecutive patients with Buruli ulcer seen between 2005 and 2011 at the Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB) in Pobe, Benin. Clinical Buruli ulcer cases were confirmed by PCR and tested for HIV infection with two serological tests. As suggested in our study,<sup>6</sup> severe Buruli ulcer was defined as presenting with oedematous, bone, large ( $\geq 15$  cm in diameter), or multifocal lesions. The effect of HIV on Buruli ulcer severity was tested by logistic regression, as implemented in the glm function of the R software. Access to the data registry was approved by the institutional review board of the CDTUB and the national Buruli ulcer control authorities.

1511 Buruli ulcer cases were included, of whom 78% ( $n=1177$ ) were PCR-confirmed. HIV testing was positive in 34 of all 1511 patients (2.3%) and in 25 of 500 patients (5.0%) aged 15–49 years. This proportion is significantly higher than is the 1.1% estimate reported for this age group in Benin in 2012 ( $p<0.0001$ ).<sup>7</sup> More than 70% ( $n=24$ ) of patients with HIV developed severe Buruli ulcer, compared with 50% ( $n=723$ ) of HIV-negative patients (odds ratio [OR] 2.77, 95% CI 1.32–6.33;  $p=0.006$ ). A focus on PCR-confirmed Buruli ulcer cases further validated this finding (OR 2.59, 1.06–7.27;  $p=0.037$ ). The effect of HIV on Buruli ulcer severity was driven mainly by an increased frequency of large or oedematous lesions in patients with HIV and Buruli ulcer. The effect of HIV on Buruli ulcer severity was driven mainly by an increased frequency of large or oedematous lesions in patients with HIV and Buruli ulcer (OR for large lesions in patients with Buruli ulcer with and without HIV 2.32, 95% CI 1.16–4.76,  $p=0.0174$ ; OR for oedematous lesions 1.93, 0.94–3.86,  $p=0.0740$ ; OR for bone lesions 0.96, 0.15–3.24,  $p=0.95$ ; and OR for multifocal lesions 1.55, 0.25–5.29,  $p=0.58$ ). Adjustment for age and sex did not modify the results.

Although HIV–Buruli ulcer co-infection is a rare clinical event, our data support O'Brien and colleagues' hypothesis of a significant effect of HIV infection on Buruli ulcer severity. We also report evidence suggestive of a higher incidence of HIV infection in patients with Buruli ulcer compared with the general population, although the absence of local controls in our design calls for further confirmation. This evidence offers insights of profound significance with regards to Buruli ulcer physiopathology, and opens new avenues for the development of novel preventive and therapeutic strategies.

We declare no competing interests. We thank staff at the Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB), Pobè, Bénin; staff of the Laboratoire de Bactériologie, CHU, Angers, France; and staff from the Institut National de la Santé et de la Recherche Médicale (INSERM) U1163 for helpful discussions; and acknowledge support from la Fondation Raoul Follereau. VBQ acknowledges support from the Fondation Imagine. ML and AA acknowledge support from the Agence Nationale de la Recherche (ANR). AA acknowledges support from the Fondation pour la Recherche Médicale (FRM, grant number DM12009117308). AA and ML acknowledge support from the INSERM.

Quentin B Vincent,  
Marie-Françoise Ardant,  
Laurent Marsollier, Annick Chauty,  
\*Alexandre Alcaïs, on behalf of the  
Franco-Beninese Buruli Research  
Group (listed in appendix)  
alexandre.alcais@inserm.fr

Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Recherche Médicale U1163 (INSERM), Paris, France (QBV, AA); Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris, France (QBV, AA); Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB), Pobè, Benin (MFA, AC); Fondation Raoul Follereau, Paris, France (MFA, AC); Institut National de la Recherche Médicale U892 (INSERM) et CNRS U6299, équipe 7, Université et CHU d'Angers, Angers, France (LM); and St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA, and Unité de Recherche Clinique, Paris Centre Descartes Necker Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France (AA)

- 1 O'Brien DP, Comte E, Serafini M, et al. The urgent need for clinical, diagnostic, and operational research for management of Buruli ulcer in Africa. *Lancet Infect Dis* 2014; **14**: 435–40.
- 2 Raghunathan PL, Whitney EAS, Asamoah K, et al. Risk factors for Buruli ulcer disease (*Mycobacterium ulcerans* infection): results from a case-control study in Ghana. *Clin Infect Dis* 2005; **40**: 1445–53.
- 3 Johnson RC, Nackers F, Glynn JR, et al. Association of HIV infection and *Mycobacterium ulcerans* disease in Benin. *AIDS* 2008; **22**: 901–03.
- 4 Toll A, Gallardo F, Ferran M, et al. Aggressive multifocal Buruli ulcer with associated osteomyelitis in an HIV-positive patient. *Clin Exp Dermatol* 2005; **30**: 649–51.
- 5 Delaporte E, Alfandari S, Piette F. *Mycobacterium ulcerans* associated with infection due to the human immunodeficiency virus. *Clin Infect Dis* 1994; **18**: 839.
- 6 Vincent QB, Ardant M-F, Adeye A, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Africa: a cohort study. *Lancet Glob Health* 2014; **2**: e422–30.
- 7 UNAIDS. Country Epidemiology: Benin. 2012. <http://www.unaids.org/en/regionscountries/countries/benin/> (accessed Aug 5, 2014).

## Research on very early ART in neonates at risk of HIV infection

The news that the Mississippi child previously in sustained viral remission now has detectable HIV<sup>1</sup> is profoundly disappointing. We write to clarify that research on the use of very early antiretroviral therapy (ART) in infants born to mothers with HIV is still important. This development, however, gives reason to think carefully about how to do this research.

Although the Mississippi case did not show how to cure HIV, it is the first documented case in which an infant infected with HIV remained free of detectable HIV for a sustained period, despite not being on ART and not being an elite controller.<sup>2</sup> Independent evidence—including animal data,<sup>3</sup> evidence from adult cohorts initiating ART in acute infection,<sup>4</sup> and data on post-exposure prophylaxis in adults<sup>5</sup>—suggests that very early combination ART (whether alone or in conjunction with other interventions) might enable infants to avoid a lifetime of ART. Nevertheless, why the Mississippi child had viral remission for more than 2 years, why the virus re-emerged, or whether modifications of this strategy could lead to longer periods of remission (which would enable avoidance of ART and its associated side effects) is not certain.

Research on the use of very early combination therapy in infants of mothers with HIV to achieve viral remission is therefore crucially important, as is the question of how to maintain potential lifelong viral remission. We believe the ethical framework in our Personal View<sup>6</sup> is still the right way to proceed with this research. The challenge lies in specification of appropriate adjustments to the scientific plan and the ethical safeguards in light of this new information. For instance, perhaps the Mississippi child did not have treatment for long enough to

achieve lifelong viral remission, but how long treatment should be given for the strategy to work is unclear. This lack of clarity suggests an even greater need for an ongoing process in which experts reach consensus on criteria for discontinuation and restarting of treatment, on the basis of this new information and other future developments. These criteria will need to be rigorous, to protect the individual children participating in research and to create the best chance of answering the question of whether very early therapy can lead to sustained viral remission of HIV.

We declare that we have no competing interests. This research was supported and sponsored by the IMPAACT HIV CURE Scientific Committee, the IMPAACT network, the Division of AIDS at the NIH, and the Intramural Research Program of the NIH in the Warren G Magnuson Clinical Center. Work on this letter by DP and HT was partly supported by the Johns Hopkins Center for AIDS Research, funded by the National Institutes of Health (grant number 1P30AI094189). The opinions expressed are the view of the authors; they do not represent any position or policy of the US National Institutes of Health, the Public Health Service, or the Department of Health and Human Services.

\*Seema K Shah, Deborah Persaud,  
David S Wendler, Holly A Taylor,  
Hannah Gay, Mariana Kruger, Christine Grady  
shahse@mail.nih.gov

NIH Department of Bioethics, 10 Center Dr,  
10/1C118, Bethesda, MD 20892, USA

- 1 National Institutes of Health News, press release: "Mississippi baby" now has detectable HIV, researchers find. July 10, 2014. <http://www.niaid.nih.gov/news/newsreleases/2014/pages/mississippibabyhiv.aspx> (accessed Aug 12, 2014).
- 2 Persaud D, Gay H, Ziemniak C, et al. Absence of detectable HIV-1 viremia after treatment cessation in an infant. *N Engl J Med* 2013; **369**: 1828–35.
- 3 Haase AT. Early events in sexual transmission of HIV and SIV and opportunities for interventions. *Ann Rev Med* 2011; **62**: 127–39.
- 4 Sáez-Cirión A, Bacchus C, Hocqueloux L, et al. Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI study. *PLoS Pathog* 2013; **9**: e1003211.
- 5 Young TN, Arens FJ, Kennedy GE, Laurie JW, Rutherford GW. Antiretroviral post-exposure prophylaxis (PEP) for occupational HIV exposure. *Cochrane Database Syst Rev* 2007; **1**: CD002835.
- 6 Shah SK, Persaud D, Wendler DS, et al. Research into a functional cure for HIV in neonates: the need for ethical foresight. *Lancet Infect Dis*; **14**: 893–98

See Personal View page 893

See Online for appendix

## Supplementary webappendix

This webappendix formed part of the original submission. We post it as supplied by the authors.

Supplement to: Vincent QB, Ardant M-F, Marsollier L, et al. HIV infection and Buruli ulcer in Africa. *Lancet Infect Dis* 2014; **14**: 796–97.

# Supplementary Appendix

Supplement to: **HIV infection and Buruli ulcer in Africa, Correspondence in the Lancet Infectious Diseases, 2014.**

Vincent QB, Ardant MF, Marsollier L, Chauty A, Alcaïs A on behalf of the Franco-Beninese Buruli Research Group

## ***The Franco-Beninese Buruli Research Group :***

### ***France:***

Paris: Quentin B. Vincent<sup>1,2</sup>, Laurent Abel<sup>1,2,4</sup>, Christian Johnson<sup>3</sup>, Alexandre Alcaïs<sup>1,2,4,5</sup>

Angers: Estelle Marion<sup>6</sup>, Marie Kempf<sup>7</sup>, Jane Cottin<sup>7</sup>, Jean-Paul Saint-André<sup>8</sup>, Laurent Marsollier<sup>6</sup>

### ***Benin:***

Pobè: Marie-Françoise Ardant<sup>3,9</sup>, Ambroise Adeye<sup>9</sup>, Aimé Goundote<sup>9</sup>, Annick Chauty<sup>3,9</sup>

Cotonou: Didier Agossadou<sup>10</sup>

(1) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Recherche Médicale U1163 (INSERM), Paris, France

(2) Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris, France

(3) Fondation Raoul Follereau, Paris, France

(4) St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA

(5) Unité de Recherche Clinique, Paris Centre Descartes Necker Cochin, Assistance Publique- Hôpitaux de Paris, Paris, France

(6) Institut National de la Recherche Médicale U892 (INSERM) et CNRS U6299, équipe 7, Université et CHU d'Angers, Angers, France

(7) Laboratoire de Bactériologie, CHU d'Angers, Angers, France

(8) Laboratoire d'Anatomie Pathologique, CHU d'Angers, Angers, France

(9) Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB), Pobè, Benin

(10) Programme de Lutte Contre la Lèpre et l'Ulcère de Buruli, Ministère de la Santé, Cotonou, Bénin

## ARTICLE 4

*Virginie Pommelet, **Quentin B Vincent**, Marie-Françoise Ardant, Ambroise Adeye,  
Anca Tanase, Laura Tondeur, Adelaide Rega, Jordi Landier, Estelle Marion, Alexandre  
Alcaïs, Laurent Marsollier, Arnaud Fontanet, and Annick Chauty*

### **Findings in Patients From Benin With Osteomyelitis and Polymerase Chain Reaction–Confirmed *Mycobacterium ulcerans* Infection**

Publié dans Clinical Infectious Diseases.

July 21, 2014, ePub ahead of Print





# Findings in Patients From Benin With Osteomyelitis and Polymerase Chain Reaction–Confirmed *Mycobacterium ulcerans* Infection

Virginie Pommelet,<sup>1</sup> Quentin B. Vincent,<sup>2,3</sup> Marie-Françoise Ardant,<sup>7</sup> Ambroise Adeye,<sup>7</sup> Anca Tanase,<sup>4</sup> Laura Tondeur,<sup>1</sup> Adelaide Rega,<sup>4</sup> Jordi Landier,<sup>1</sup> Estelle Marion,<sup>7,8,9</sup> Alexandre Alcaïs,<sup>2,3,5,6</sup> Laurent Marsollier,<sup>8,9</sup> Arnaud Fontanet,<sup>1,6</sup> and Annick Chauty<sup>7</sup>

<sup>1</sup>Emerging Diseases Epidemiology Unit, Institut Pasteur, <sup>2</sup>Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Recherche Médicale U980 (INSERM), <sup>3</sup>Université Paris Descartes, Sorbonne Paris Cité, Imagine Institute, <sup>4</sup>Department of Pediatric Radiology, Robert Debré Children University Hospital, Assistance Publique-Hôpitaux de Paris, <sup>5</sup>CIC-0109 Cochin-Necker Inserm, Unité de Recherche Clinique, Paris Centre Descartes Necker Cochin, Assistance Publique-Hôpitaux de Paris et EA 3620, Université Paris Descartes, and <sup>6</sup>Conservatoire National des Arts et Métiers, Paris; <sup>7</sup>Centre de Diagnostic et de Traitement de la Lèpre et de l'Ulçère de Buruli, Fondation Raoul Follereau, Pobè, Bénin; and <sup>8</sup>ATOMycA, Inserm Avenir Team, CRCNA, Inserm U892, 6299 CNRS, Université et CHU, and <sup>9</sup>LUNAM, Université d'Angers, France

**Background.** *Mycobacterium ulcerans* is known to cause Buruli ulcer (BU), a necrotizing skin disease leading to extensive cutaneous and subcutaneous destruction and functional limitations. However, *M. ulcerans* infections are not limited to skin, and osteomyelitis, still poorly described in the literature, occurs in numerous young patients in Africa.

**Methods.** In a retrospective matched case-control study conducted in a highly endemic area in Benin, we analyzed demographic, clinical, biological, and radiological features in all patients with *M. ulcerans* infections with bone involvement, identified from a cohort of 1257 patients with polymerase chain reaction–proved *M. ulcerans* infections.

**Results.** The 81 patients studied had a median age of 11 years (interquartile range, 7–16 years) and were predominantly male (male–female ratio, 2:1). Osteomyelitis was observed beneath active BU lesions (60.5%) or at a distance from active or apparently healed BU lesions (14.8%) but also in patients without a history of BU skin lesions (24.7%). These lesions had an insidious course, with nonspecific clinical findings leading to delayed diagnosis. A comparison with findings in 243 age- and sex-matched patients with BU without osteomyelitis showed that case patients were less likely to have received BCG immunization than controls (33.3% vs 52.7%;  $P = .01$ ). They were also at higher risk of longer hospital stay (118 vs 69 days;  $P = .001$ ), surgery (92.6% vs 63.0%;  $P = .001$ ), and long-term crippling sequelae (55.6% vs 15.2%;  $P < .001$ ).

**Conclusions.** This study highlighted the difficulties associated with diagnosis of *M. ulcerans* osteomyelitis, with one-fourth of patients having no apparent history of BU skin lesions, including during the current course of illness. Delays in treatment contributed to the high proportion (55.6%) of patients with crippling sequelae.

**Keywords.** Buruli ulcer; osteomyelitis; *Mycobacterium ulcerans*.

Buruli ulcer (BU) caused by *Mycobacterium ulcerans* is with tuberculosis and leprosy among the most common human mycobacterial diseases. Children and adolescents bear the highest burden of the disease [1]. In

1998, The World Health Organization (WHO) recognized it as an emerging infectious disease in West and Central Africa with an important public health impact and classified it as a neglected tropical disease [2]. Environmental and ecological studies show that the disease is most common in populations living near rivers, swamps and wetlands [3–5], and the role of living agents such as aquatic insects or mosquitoes as reservoirs and vectors of *M. ulcerans* has been proposed but remains controversial [6–11].

Classically, BU is a necrotizing skin disease characterized by an indolent painless course leading to

Received 7 April 2014; accepted 9 July 2014.

Correspondence: Arnaud Fontanet, MD, PhD, Institut Pasteur, Emerging Diseases Epidemiology Unit, 75015 Paris, France (fontanet@pasteur.fr).

## Clinical Infectious Diseases

© The Author 2014. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.

DOI: 10.1093/cid/ciu584

extensive cutaneous and subcutaneous destruction, functional limitations and disabilities. It has a large spectrum of clinical presentations, ranging from painless subcutaneous nodules to extensively undermined ulcers and disseminated forms, related partly to a delays in seeking medical-care [12, 13]. The lesions of BU are mainly due to the cytotoxic and immunosuppressive properties of the toxin secreted by *M. ulcerans*, known as mycolactone [14, 15]. Lesions are classified according to the clinical presentation—nonulcerative (nodule, plaque, edema) or ulcerative—and according to the cross-sectional diameter of the lesion [16].

Until recently, treatment recommendations focused on extensive excision of the lesion and skin grafting leading to important functional and esthetic sequelae. Significant progress has been made with the demonstration of the efficacy of an 8-week course of rifampicin plus streptomycin chemotherapy [16]; more recently, the efficacy of oral treatment with clarithromycin and rifampicin has been demonstrated in mild forms of BU [17, 18].

However, *M. ulcerans* infections are not limited to skin. Osteomyelitis has been described in patients with BU in Africa, with rates ranging from 6% [1] to more than 14% [13, 19] and with young boys at increased risk [20]. Bone invasion was documented both beneath skin lesions (contiguous osteomyelitis) and also remote from them (metastatic osteomyelitis) [21–23]. Meanwhile, osteomyelitis and extensive forms have remained poorly described in the literature and excluded from most of the clinical trials. Recommendations concerning the diagnosis and management of these severe forms are scarce, and the conditions may lead to crippling disabilities and amputations [21].

We recently published a comprehensive description and analysis of BU clinical epidemiology based on laboratory-confirmed cases [20]. In this article, we focus on BU osteomyelitis and describe in detail the demographics, clinical presentation, biological and radiological features, management, and outcome in patients with *M. ulcerans* osteomyelitis.

## PATIENTS AND METHODS

### Study Design and Definitions

This retrospective matched case-control study was conducted in the Centre de Diagnostic et de Traitement de la Lèpre et de l'Ul-cère de Buruli (CDTLUB) in Pobè, Benin. The CDTLUB has treated 1819 BU cases from 2003 to 2011, of which 1257 were confirmed by polymerase chain reaction (PCR).

We reviewed all patients with PCR-confirmed *M. ulcerans* infection and bone involvement admitted to the CDTLUB from 2003 to December 2011. A case patient was defined as one with clinical and radiological osteomyelitis and laboratory-confirmed *M. ulcerans* infection. The clinical presentation of

osteomyelitis has a large spectrum, ranging from an open wound exposing fractured bone to an indolent draining fistula to local swelling and tenderness at clinical examination; physicians made the diagnosis according to their own experience [24–26]. Controls were defined as patients with PCR-confirmed BU and no bone lesions admitted to the CDTLUB during the same period. Three controls, matched for both age and sex, were randomly assigned to each case patient. Case patients and controls were identified through the CDTLUB database, prospectively created along with the notification of *M. ulcerans* infections to the WHO [20].

### Data Collection

Detailed patient data regarding demographic characteristics, medical history, course of present illness, clinical and radiological features, microbiological findings, management (medical and surgical treatment), and outcome were collected from the medical and laboratory records of the CDTLUB in Pobè, including the WHO BU.01 form. The BCG immunization status was established by the presence or the absence of a scar on the left forearm. Additional data regarding PCR analysis were collected from the records of the microbiological laboratory of Angers, France. When available, we reviewed digital photographs of the skin lesions of the case patients at the time of admission and during follow-up. All radiographs were reviewed by experienced radiologists in France (A. R. and A. T.). Data were collected using EpiData software (version 3.0; The EpiData Association).

### Laboratory Methods

Biological samples (swab, fine needle aspiration, and skin and bone biopsy) were collected and handled according to WHO guidelines [27]. *M. ulcerans* infection was confirmed by IS2404 PCR analysis for all patients included in the study. Routine laboratory tests (complete blood count, erythrocyte sedimentation rate and human immunodeficiency virus testing) were performed in the laboratory of the CDTLUB. Radiographs were obtained in local hospitals.

### Statistical Analysis

Continuous variables were summarized using medians and interquartile ranges (IQRs), and categorical variables were summarized using frequencies and percentages with their 95% confidence intervals (CIs). The data consisted of individually matched sets with 1 case patient and 3 controls. *P* values for the comparisons of categorical variables between case patients and controls were obtained using conditional logistic regression analysis. Matched sets were dropped from the analysis if no data were available for the case patient or for all 3 controls. Continuous variables (eg, hemoglobin level) were compared using linear regression analysis, after log transformation when not

normally distributed, adjusted for age and sex. Variables were compared among subgroups of case patients using Fisher exact, Mann–Whitney and Kruskal–Wallis tests, where appropriate. Statistical significance was set at the 5% level (2-sided *P* value). Analyses were conducted using Stata software (version 12.0; StataCorp).

## RESULTS

### Description of Patients With Osteomyelitis

From 2003 to 2011, a total of 80 patients with PCR-confirmed BU with osteomyelitis were identified from the database of the CDTLUB. They were age- and sex-matched with 240 controls without osteomyelitis. Of these 80 case patients, 2 were excluded after review of the paper files because of insufficient documentation of osteomyelitis, and 4 were considered not to have osteomyelitis; a total of 74 patients with osteomyelitis (5.9%; 95% CI, 4.6%–7.2%) were thus identified from the database. Of the 240 controls, 9 (3.7%; 95% CI, 5.4%–12.6%) were found to have bone lesions and were reclassified as case patients with osteomyelitis. New age- and sex-matched controls were randomly assigned to these 9 new case patients. Two patients with BU with osteomyelitis also had malum perforans pedis related to leprosy and were excluded from the case-control analysis. Altogether, 81 patients with osteomyelitis and their 243 controls were kept for the analysis presented herein.

The case patients had a median (IQR) age of 11 (7–16) years and were predominantly male (male-female ratio, 2:1). The median (IQR) time from symptom onset to the first hospital visit was 40 (12–52) weeks, and 19 (23.7%) patients waited for more than a year. All of them (100%) had a complex treatment history, including traditional healing (91.4%), inappropriate antibiotics (38.3%), or inappropriate minor surgery (40.7%) (Table 1). At the first visit, 9 patients (11.1%) were malnourished, 9 (11.1%) had a history of fever, and 32 (39.5%) had local pain. Osteomyelitis was located on the lower limbs in 51 patients (63.0%). Regional lymphadenopathy was documented in 36 (44.4%) patients, and limited function and movement in 71 (87.6%) (Table 1).

Overall, 49 osteomyelitis lesions (60.5%) occurred beneath typical BU skin lesions, and these patients were considered to have *contiguous osteomyelitis*. In 4 instances, the BU skin lesion was no longer active, but the scar and the history were typical of past BU. The cross-sectional diameter of the BU lesions exhibited by patients with contiguous osteomyelitis was larger than that in controls, with lesions >15 cm in 27 case patients (61.4%) versus 41 controls (28.9%;  $P < .001$ ). Twelve patients (14.8%) had bone lesions distant from either an active ( $n = 6$ ) or a past ( $n = 6$ ) BU skin lesion, and were considered to have *metastatic osteomyelitis*, and 20 patients (24.7%) had no visible active or past BU lesion. The skin

lesions observed next to the bone lesions in these patients were comparable to those observed in nonspecific osteomyelitis. These patients were considered to have *exclusive osteomyelitis*. Table 2 shows characteristics of the skin lesions contiguous to osteomyelitis documented at the first visit in these 3 categories of patients.

All case patients included in the study had *M. ulcerans* infection confirmed with PCR, with PCR testing of a bone biopsy specimen in 31 case patients (43.1%), of pus from the fistula in 5 (6.9%), and of skin biopsy, fine needle aspiration, or swab specimens of the skin lesion contiguous to the bone infection in 36 (50.0%); the origin of the positive PCR specimen was not specified in 9 patients (Supplementary Table 1). For the 20 case patients categorized as having exclusive osteomyelitis, the diagnosis of *M. ulcerans* infection was based on PCR testing of pus from the fistula ( $n = 2$ ), bone biopsy specimens ( $n = 10$ ), or swab specimens of nonspecific skin lesions ( $n = 8$ ).

The diagnosis of osteomyelitis with radiography preceded initiation of treatment in 53 patients (63.4%). For the remaining 28 patients, radiographic diagnosis took place a median (IQR) of 25.5 (3–56.5) days after initiation of treatment. The vast majority of these patients with delayed diagnosis ( $n = 24$ ; 86%) had contiguous osteomyelitis, and 4 had metastatic ( $n = 3$ ) or exclusive ( $n = 1$ ) osteomyelitis. Seventeen patients (21.0%) had multiple bone lesions. Metastatic bone lesions appeared after antibiotics were started in 5 patients, and the median time from the first day of treatment to the diagnosis of the first metastatic bone lesion was 88 days. A total of 99 radiographs from 76 patients were reviewed (Supplementary Table 2); they revealed features of chronic (75.8%) and extensive (56.6%, affecting >30% of the bone surface) lesions, located predominantly on long bones and inferior limbs (Figure 1).

### Case-Control Analysis

Table 1 shows demographic characteristics, underlying conditions, and history of the present illness in the 81 case patients with osteomyelitis and 243 age- and sex-matched control subjects included in the analysis. Human immunodeficiency virus (2 controls vs no case patients) was not associated with bone lesions in this cohort. Case patients were less likely to have received BCG immunization than controls (33.3% vs 52.7%;  $P = .01$ ) and more likely to have a history of BU (14.8% vs 2.1%;  $P = .001$ ). Patients with bone lesions had waited longer before coming to the CDTLUB or affiliated primary healthcare centers ( $P = .001$ ) and during that time were more likely to have sought traditional healing or received inappropriate antibiotics or minor surgical interventions ( $P = .001$ ).

Clinical features of the patients and descriptions of the main lesions are presented in Table 1. Patients with bone lesions controls were more likely than controls to have functional limitations (87.6% vs 45.7%, respectively;  $P < .001$ ) and multifocal

**Table 1. Demographic Characteristics, Medical History, Course of Present Illness, and Clinical Features in Case Patients and Matched Controls**

Characteristic	Patients, No. (%) <sup>a</sup>		P Value
	Case Patients (n = 81)	Controls (n = 243)	
Age, median (IQR), y	11 (7–16)	11 (7–15)	. . .
Sex			
Female	27 (33.3)	81 (33.3)	. . .
Male	54 (66.7)	162 (66.7)	
Medical history			
BCG vaccination <sup>b</sup>	27 (33.3)	128 (52.7)	.01 <sup>c</sup>
Family history of BU	6 (7.4)	28 (11.5)	.28
Personal history of BU <sup>d</sup>	12 (14.8)	5 (2.1)	<.001
HIV infection	0	2 (0.8)	NS
Time from onset of present illness to diagnosis			
≤1 mo	11 (13.7)	91 (37.8)	<.001 <sup>e</sup>
1–6 mo	23 (28.7)	124 (51.4)	
>6 mo	46 (57.5)	26 (10.8)	
History of injury	22 (27.2)	36 (14.8)	.01
Initial presentation of lesion as described by patient			
Edema	27 (35.5)	57 (24.3)	<.001 <sup>f</sup>
Nodule	8 (10.5)	81 (34.5)	
Tumefaction	19 (23.5)	67 (27.7)	
Ulcer	12 (14.8)	25 (10.3)	
Other <sup>g</sup>	10 (13.2)	5 (2.1)	
Management before first hospital visit			
Traditional healing	74 (91.4)	165 (67.9)	<.001
Antibiotics (oral or IV)	31 (38.3)	41 (16.9)	<.001
Topical antibiotics	34 (42.0)	25 (10.3)	<.001
Minor surgery	33 (40.7)	15 (6.2)	<.001
Clinical signs			
Fever	9 (11.1)	21 (8.6)	.50
Regional lymphadenopathy	36 (44.4)	89 (36.6)	.21
Functional limitation	71 (87.6)	111 (45.7)	<.001
Multifocal lesions	27 (33.3)	7 (2.9)	<.001
Location of main lesion			
Upper limbs	27 (33.3)	91 (37.4)	.03
Lower limbs	51 (63.0)	121 (49.8)	
Abdomen	1 (1.2)	12 (4.9)	
Thorax	2 (2.5)	9 (3.7)	
Other <sup>g</sup>	0	10 (4.1)	

Abbreviations: BCG, Bacillus Calmette-Guerin; BU, Buruli ulcer; HIV, human immunodeficiency virus; IQR, interquartile range; IV, intravenous; NS, not significant.

<sup>a</sup> Data represent No. (%) of patients or controls unless otherwise noted.

<sup>b</sup> A history of BCG vaccination was determined by the presence of a BCG vaccine scar (on the left arm).

<sup>c</sup> Twenty of 81 matched groups were dropped from the conditional logistic regression analysis because of missing data.

<sup>d</sup> Defined as the presence of a stellate scar suggestive of past BU or medical history of a prior BU.

<sup>e</sup> One of 81 matched groups was dropped from the conditional logistic regression analysis because of missing data.

<sup>f</sup> Five of 81 matched groups were dropped from the conditional logistic regression analysis because of missing data.

<sup>g</sup> Abscess, fistula, back, buttocks and perineum, head, and neck.

lesions (33.3% vs 2.9%;  $P < .001$ ). The *M. ulcerans* infection was confirmed in all case patients and controls by PCR of a swab, fine needle aspiration, or biopsy (skin or bone) specimen

(Supplementary Table 1). The median time to biological confirmation was longer for case patients than for controls (7 vs 1 days, respectively;  $P = .001$ ).

**Table 2. Characteristics Associated With Skin Lesions Contiguous to Bone Lesions Exhibited at First Hospital Visit**

Characteristic	Skin Lesions, No. (%) <sup>a</sup>	Lesions Not Suggestive of BU (n = 32)	
	Typical BU Lesions (Contiguous Osteomyelitis; n = 49)	Past or Active Distant BU Lesion (Metastatic Osteomyelitis; n = 12)	No Past or Active BU Lesion (Exclusive Osteomyelitis; n = 20)
<b>Cutaneous lesions</b>			
Ulcer	39 (79.6)		
Edema	23 (46.9)	4 (33.3)	5 (25.0)
Plaque	23 (46.9)		1 (5.0)
Nodule	0		
Scar	4 (8.3)		
Tumefaction		9 (75.0)	20 (100.0)
Fistula	18 (36.7)	9 (75.0)	19 (95.0)
Visible bone	13 (26.5)	2 (16.7)	3 (15.0)
<b>Size</b>			
<5 cm	6 (13.6)		
5–15 cm	11 (25.0)		
>15 cm	27 (61.4)		
MD	5 (10.4)		
Age, median (IQR), y	12 (8–16)	9 (6.5–16.5)	9.5 (6.5–14)
Time from onset of symptoms to first visit, median (IQR), d	36 (12–52)	52 (12–208)	32 (12–52)
BCG vaccination <sup>b</sup>	14 (28.5)	3 (25.6)	10 (50.0)
BCG status unknown	15 (30.6)	1 (8.3)	4 (20.0)
Multifocal bone lesions	9 (18.4)	5 (41.7)	4 (20.0)
<b>PCR confirmation of <i>M. ulcerans</i></b>			
Bone biopsy	13 (31.7)	8 (72.7)	10 (50.0)
Pus from fistula	2 (4.1)	1 (8.3)	2 (10.0)
Specimen from skin lesion <sup>c</sup>	26 (53.1)	2 (16.7)	8 (40.0)
Not specified	8 (16.3)	1 (8.3)	0

Abbreviations: BCG, Bacillus Calmette-Guerin; BU, Buruli ulcer; IQR, interquartile range; MD, missing data; *M. ulcerans*, *Mycobacterium ulcerans*; PCR, polymerase chain reaction.

<sup>a</sup> Data represent No. (%) of lesions unless otherwise noted.

<sup>b</sup> A history of BCG vaccination was determined by the presence of a BCG vaccine scar (on the left arm).

<sup>c</sup> Swab, fine needle aspiration, or skin biopsy samples of lesions contiguous to the bone infection.

Routine laboratory tests showed that patients with osteomyelitis were more likely than controls to present with anemia and elevated white blood cell counts (Supplementary Table 3). The other hematological features were similar in both groups.

All patients were managed according to the WHO recommendations with adapted wound care and physiotherapy to prevent disability. Case patients were more likely than controls to be hospitalized (96.3% vs 64.2%;  $P = .001$ ), and for a longer median duration (118 vs 69 days;  $P = .001$ ). The management of patients with osteomyelitis was more complex; they received more frequently additional antibiotics (44.4% vs 11.5%;  $P = .001$ ) and underwent surgery more frequently (92.6% vs 63.0%;  $P = .001$ ) (Table 3, Figure 1, and Supplementary Figure 1).

Patients with bone lesions were less likely to recover from functional limitations than those with no bone lesions (28.8%

vs 72.1%;  $P = .001$ ). Treatment failure—defined as the absence of healing at a long-term follow-up visit in a patient who received appropriate medical treatment and/or underwent surgery—was more common in case patients than in controls (7.9% vs 1.6%;  $P = .05$ ).

## DISCUSSION

This is the first systematic retrospective analysis of bone lesions in a large cohort of 1257 PCR-confirmed *M. ulcerans* infections. The analysis in our 81 case patients shows that osteomyelitis can be observed beneath active BU lesions or at a distance from active or apparently healed BU lesions but also in patients without any apparent history of BU skin lesions. These lesions have an insidious course with nonspecific clinical features, which often





**Figure 1.** Clinical and radiological evolution during 1 year of chronic *Mycobacterium ulcerans* osteomyelitis of the right tibia in a 10-year old boy. *A*, Buruli ulcer of the right leg with contiguous tibia osteomyelitis; the radiograph shows a sequestrum with destruction of the metaphysis. *B*, *C*, Clinical and radiological evolution at 3 and 10 months respectively. The patient received an 8-week antibiotic regimen (streptomycin and rifampicin) and underwent surgery (skin and bone debridement with sequestrectomy).

leads to delayed diagnosis. The management of patients with bone lesions is complex; they tend to have longer hospital stays and higher risks of surgery and long-term crippling sequelae than patients with BU without osteomyelitis.

It is not clear how *M. ulcerans* infects bones. Indeed, bone localization and metastatic spread of *M. ulcerans* are surprising, considering that this agent grows optimally at 30°–33°C. Certain isolates of *M. ulcerans* seem to be more thermotolerant [28], and strains recovered from bone lesions in this study will be sequenced for comparison with strains obtained from skin lesions. As described in other studies, contiguous bone infection develops by direct spread of *M. ulcerans* from an overlying BU skin lesion that is most frequently large and ulcerative [21]. Minor trauma has also been suggested as a risk factor for non-specific osteomyelitis, and patients with osteomyelitis in our study were more likely than controls to have experienced recent trauma. Microtrauma in the bone medulla could facilitate the local infection. Metastatic *M. ulcerans* osteomyelitis—bone infections developing remotely from an active or apparently healed BU, as metastatic skin lesions—are believed to result from lymphohematogenous dissemination of the mycobacteria from skin lesions. The fact that bone lesions were frequently observed on the metaphysis of bones reinforces this hypothesis. Also coinciding with this hypothesis, patients with bone infection were more likely to have multifocal lesions. Thus, once the bacteria spreads in the bloodstream, it is likely to disseminate to multiple sites. Therefore, BU should be considered a potentially systemic disease, and clinicians should search for other sites once a single focus is identified [12, 22, 23, 29].

In this study, in addition to these known presentations of *M. ulcerans* osteomyelitis, we identified patients (25% of all patients with osteomyelitis) who had no evidence of past or present BU skin lesions. This seems to be a third category of *M. ulcerans* osteomyelitis, which we called *exclusive osteomyelitis*. The skin lesions of these patients were those of nonspecific osteomyelitis (eg, tumefaction, edema, and fistula) [25, 30], and had none of the characteristics of BU skin lesions (eg, painless nodule, plaque or undermined ulcer) [16]. These case patients may have been in contact with *M. ulcerans* in the past and the primary infection may have gone unnoticed. The delayed appearance of a bone lesion would then result from reactivation of a latent infection, as was also suggested for the development of bone lesions beneath or distant from healed BU lesions [31]. Although the diagnosis of *M. ulcerans* infection may be missed in patients with exclusive osteomyelitis owing to the absence of BU lesions, the patients in this study with *M. ulcerans* osteomyelitis had similar delays to diagnosis, radiographic bone lesions, response to treatment, and outcomes regardless of the osteomyelitis form, exclusive, contiguous, or metastatic.

Overall, *M. ulcerans* osteomyelitis has a long-standing indolent course and therefore shares common demographic

**Table 3. Management and Outcome in Case Patients and Matched Controls**

Management and Outcome	Patients, No. (%) <sup>a</sup>		P Value
	Case Patients (n = 81)	Controls (n = 243)	
Hospitalization	78 (96.3)	156 (64.2)	<.001
Duration, median (IQR), d	118 (73–184)	69 (48–103)	<.001
Therapeutic intervention			
Only antibiotics	5 (6.2)	86 (35.4)	<.001
Antibiotics and surgery	75 (92.6)	153 (67.1)	<.001
Antibiotic regimen			
S/R	78 (96.3)	231 (95.1)	.75
C/R	2 (2.5)	6 (2.5)	
None	1 (1.2)	4 (1.6)	
Duration, median (IQR), d	55 (55–56)	55 (54–56)	
Treatment withdrawal	1 (1.2)	22 (9.1)	.04
Additional antibiotic	36 (44.4)	28 (11.5)	.001
Cloxacillin	27 (33.3)	20 (8.2)	
Ciprofloxacin	11 (13.6)	1 (0.4)	
Duration, median (IQR), d	15 (10–21)	10 (8–10)	NA
Surgery	75 (92.6)	153 (63.0)	<.001
Number of surgical procedures			
1	12 (16.0)	7 (4.6)	.008
2	23 (30.7)	57 (37.2)	
≥3	40 (53.3)	89 (58.2)	
Type of surgery			
Debridement	70 (86.4)	150 (61.7)	<.001
Skin grafting	30 (37.0)	145 (59.7)	<.001
Sequestrectomy	39 (48.1)	0	<.001 <sup>b</sup>
Amputation	13 (16.0)	0	<.001 <sup>b</sup>
End of treatment assessment			
Healed without functional limitations	5 (6.2)	32 (13.2)	.004 <sup>c</sup>
Healed with functional limitations	7 (8.6)	3 (1.2)	
Not healed	67 (82.7)	189 (77.8)	
Lost to follow-up	2 (2.5)	19 (7.8)	
Assessment at hospital discharge			
Healed without functional limitations	16 (19.7)	66 (27.2)	.002 <sup>d</sup>
Healed with functional limitations	41 (50.6)	44 (18.1)	
Not healed	22 (27.2)	45 (18.5)	
Not assessed <sup>e</sup>	2 (2.5)	88 (36.2)	
Median healing time, days (IQR)	118 (71–198)	91 (64–130)	.003
MD	17 (21.0)	67 (27.6)	
Long-term follow-up			
Time to long-term assessment, median (IQR), d	418 (348–609)	395 (364–476)	.06

*Table 3 continued.*

Management and Outcome	Patients, No. (%) <sup>a</sup>		P Value
	Case Patients (n = 81)	Controls (n = 243)	
Healed without functional limitations	22 (27.2)	158 (65.0)	<.001 <sup>f</sup>
Healed with functional limitations	37 (45.7)	27 (11.1)	
Not healed	8 (9.9)	10 (4.1)	
Lost to follow-up	14 (17.3)	48 (19.7)	
Complications			
Death <sup>g</sup>	1	2	
Malignant lesions <sup>h</sup>	2	1	
Tetanus <sup>i</sup>	0	2	

Abbreviations: C/R, clarithromycin/rifampicin; IQR, interquartile range; MD, missing data; NA, not applicable; S/R, streptomycin/rifampicin.

<sup>a</sup> Data represent No. (%) of lesions unless otherwise noted.

<sup>b</sup> Exact logistic regression adjusted for age and sex.

<sup>c</sup> Two matched groups were dropped from the conditional logistic regression analysis because of missing data.

<sup>d</sup> Four matched groups were dropped from the conditional logistic regression analysis because of missing data.

<sup>e</sup> Two case patients (2.5%) and 83 controls (34.2%) were not hospitalized.

<sup>f</sup> Fourteen matched groups were dropped from the conditional logistic regression analysis because of missing data.

<sup>g</sup> One case patient died of a metastatic epidermoid carcinoma 18 months after initiation of chemotherapy, 1 control died of febrile seizures shortly after admission, and 1 control died 1 year after treatment of an undetermined disease.

<sup>h</sup> A metastatic epidermoid carcinoma developed in 1 case patient 7 months after initiation of treatment, a fibrosarcoma was diagnosed at admission in 1 case patient, and an osteosarcoma was diagnosed 7 months after initiation of treatment in 1 control.

<sup>i</sup> Both tetanus cases at the time of hospital admission occurred in 6-year-old boys.

(age and sex ratio), clinical, and radiological features with osteomyelitis linked to more frequent pathogens that are a major healthcare burden in low-income countries [25, 30, 32, 33]. Thus, a high index of suspicion for *M. ulcerans* should be maintained for all patients with subacute or chronic osteomyelitis in highly BU-endemic areas.

In concordance with previous findings, patients with bone infections had long and complex health-seeking itineraries and most of them combined elements of both biomedical and traditional therapeutic approaches. Buruli ulcer mainly affects rural populations with both limited geographic and financial access to health services and strong mystical beliefs. This combination of factors results in delayed medical care seeking until advanced clinical forms of the disease occur, including extensive skin destruction, multifocal lesions, and bone lesions [13, 34]. Another consequence of delayed treatment is the development of dead bone in skeletal lesions. This avascular tissue makes antibiotics essentially ineffective, and radical



debridement down to living bone is therefore essential for healing [30]. Moreover, *M. ulcerans* osteomyelitis affects young patients who are at risk of late complications caused by the growth and developmental disturbances of the affected bone. Involvement of the epiphyseal growth cartilage incurs a high risk of subsequent growth deformity. Early identification from radiographs, appropriate antimicrobial treatment, proper debridement of the infected bone, and clinical and radiological follow-up should limit the burden of disease in this specific population [26].

Fewer case patients than controls had a history of BCG vaccination. This is consistent with previous studies showing that BCG vaccination conferred some protection against severe forms of BU, such as osteomyelitis [19, 35], but induced only transient protective immunity against the development of skin lesions [36]. The prophylactic effect of BCG on severe and disseminated forms of BU has not been clearly explained so far but is reminiscent of the protective effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis [37].

In conclusion, this study allowed the identification of a third form of *M. ulcerans* osteomyelitis in patients with no evidence of active or past BU disease, which we called *exclusive osteomyelitis*. Physicians working in BU-endemic areas should be aware of the existence of this form of disease and actively search for *M. ulcerans* confirmation when dealing with patients with osteomyelitis, even in the absence of BU skin lesions. We also highlight the importance of early identification of bone lesions leading to appropriate antibiomic treatment, surgery, adapted physiotherapy, and clinical and radiological follow-up to prevent infection dissemination and extensive skin and bone destruction and therefore reduce functional limitation and disability.

## Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online (<http://cid.oxfordjournals.org>). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

## Notes

**Financial support.** This study received support from the Fondation Raoul Follereau and the French National Research Agency (grant ANR 11 CEPL 007 04 EXTRA-MU).

**Potential conflicts of interest.** All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

- Sopoh GE, Johnson RC, Chauty A, et al. Buruli ulcer surveillance, Benin, 2003–2005. *Emerg Infect Dis* **2007**; *13*:1374–6.
- World Health Organization. Buruli ulcer. Available at: <http://www.who.int/buruli/en/>. Accessed 15 July 2013.
- Uganda Buruli Group. Epidemiology of *Mycobacterium ulcerans* infection (Buruli ulcer) at Kinyara, Uganda. *Trans R Soc Trop Med Hyg* **1971**; *65*:763–75.
- Raghunathan PL, Whitney EA, Asamo K, et al. Risk factors for Buruli ulcer disease (*Mycobacterium ulcerans* infection): results from a case-control study in Ghana. *Clin Infect Dis* **2005**; *40*:1445–53.
- Johnson RC, Makoutode M, Sopoh GE, et al. Buruli ulcer distribution in Benin. *Emerg Infect Dis* **2005**; *11*:500–1.
- Portaels F, Elsen P, Guimaraes-Peres A, Fonteyne PA, Meyers WM. Insects in the transmission of *Mycobacterium ulcerans* infection. *Lancet* **1999**; *353*:986.
- Merritt RW, Walker ED, Small PL, et al. Ecology and transmission of Buruli ulcer disease: a systematic review. *PLoS Negl Trop Dis* **2010**; *4*:e911.
- Marion E, Eyangoh S, Yeramian E, et al. Seasonal and regional dynamics of *M. ulcerans* transmission in environmental context: deciphering the role of water bugs as hosts and vectors. *PLoS Negl Trop Dis* **2010**; *4*:e731.
- Marsollier L, Robert R, Aubry J, et al. Aquatic insects as a vector for *Mycobacterium ulcerans*. *Appl Environ Microbiol* **2002**; *68*:4623–8.
- Johnson PD, Azuolas J, Lavender CJ, et al. *Mycobacterium ulcerans* in mosquitoes captured during outbreak of Buruli ulcer, southeastern Australia. *Emerg Infect Dis* **2007**; *13*:1653–60.
- Fyfe JA, Lavender CJ, Handasyde KA, et al. A major role for mammals in the ecology of *Mycobacterium ulcerans*. *PLoS Negl Trop Dis* **2010**; *4*:e791.
- Walsh DS, Portaels F, Meyers WM. Buruli ulcer (*Mycobacterium ulcerans* infection). *Trans R Soc Trop Med Hyg* **2008**; *102*:969–78.
- Debacker M, Aguiar J, Steunou C, et al. *Mycobacterium ulcerans* disease (Buruli ulcer) in rural hospital, Southern Benin, 1997–2001. *Emerg Infect Dis* **2004**; *10*:1391–8.
- van der Werf TS, Stinear T, Stienstra Y, van der Graaf WT, Small PL. Mycolactones and *Mycobacterium ulcerans* disease. *Lancet* **2003**; *362*:1062–4.
- Hong H, Demangel C, Pidot SJ, Leadlay PF, Stinear T. Mycolactones: immunosuppressive and cytotoxic polyketides produced by aquatic mycobacteria. *Nat Prod Rep* **2008**; *25*:447–54.
- World Health Organization. Treatment of *Mycobacterium ulcerans* disease (Buruli ulcer): guidance for health workers. **2012**. Available at: [http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/77771/1/9789241503402_eng.pdf). Accessed 4 August 2014.
- Chauty A, Ardant MF, Marsollier L, et al. Oral treatment for *Mycobacterium ulcerans* infection: results from a pilot study in Benin. *Clin Infect Dis* **2010**; *52*:94–6.
- Nienhuis WA, Stienstra Y, Thompson WA, et al. Antimicrobial treatment for early, limited *Mycobacterium ulcerans* infection: a randomised controlled trial. *Lancet* **2010**; *375*:664–72.
- Portaels F, Aguiar J, Debacker M, et al. *Mycobacterium bovis* BCG vaccination as prophylaxis against *Mycobacterium ulcerans* osteomyelitis in Buruli ulcer disease. *Infect Immun* **2004**; *72*:62–5.
- Vincent QB, Ardant M-F, Adeye A, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Global Health* **2014**; *2*:e422–30.
- Lagarigue V, Portaels F, Meyers WM, Aguiar J. Buruli ulcer: risk of bone involvement! apropos of 33 cases observed in Benin. *Med Trop (Mars)* **2000**; *60*:262–6. [in French].
- Pszolla N, Sarkar MR, Strecker W, et al. Buruli ulcer: a systemic disease. *Clin Infect Dis* **2003**; *37*:e78–82.
- Toll A, Gallardo F, Ferran M, et al. Aggressive multifocal Buruli ulcer with associated osteomyelitis in an HIV-positive patient. *Clin Exp Dermatol* **2005**; *30*:649–51.
- Lewis JA, Miller DR, Davies SG. Osteomyelitis complicating three types of traumatic hand wound. *J Wound Care* **2004**; *13*:281–3.
- Mantero E, Carbone M, Calevo MG, Boero S. Diagnosis and treatment of pediatric chronic osteomyelitis in developing countries:

- prospective study of 96 patients treated in Kenya. *Musculoskelet Surg* **2011**; 95:13–8.
26. van Schuppen J, van Doorn MM, van Rijn RR. Childhood osteomyelitis: imaging characteristics. *Insights Imaging* **2012**; 3:519–33.
  27. World Health Organization. Guidance on sampling techniques for laboratory-confirmation of *Mycobacterium ulcerans* infection. **2010**. Available at: [http://www.who.int/buruli/Guidance\\_sampling\\_techniques\\_MU\\_infection.pdf](http://www.who.int/buruli/Guidance_sampling_techniques_MU_infection.pdf). Accessed 15 July 2013.
  28. Eddyani M, Portaels F. Survival of *Mycobacterium ulcerans* at 37 degrees C. *Clin Microbiol Infect* **2007**; 13:1033–5.
  29. Wansbrough-Jones M, Phillips R. Buruli ulcer: emerging from obscurity. *Lancet* **2006**; 367:1849–58.
  30. Lew DP, Waldvogel FA. Osteomyelitis. *Lancet* **2004**; 364:369–79.
  31. Portaels F, Silva MT, Meyers WM. Buruli ulcer. *Clin Dermatol* **2009**; 27:291–305.
  32. Peltola H, Paakkonen M. Acute osteomyelitis in children. *N Engl J Med* **2014**; 370:352–60.
  33. Stanley CM, Rutherford GW, Morshed S, Coughlin RR, Beyeza T. Estimating the healthcare burden of osteomyelitis in Uganda. *Trans R Soc Trop Med Hyg* **2010**; 104:139–42.
  34. Peeters Grietens K, Toomer E, Um Boock A, et al. What role do traditional beliefs play in treatment seeking and delay for Buruli ulcer disease? insights from a mixed methods study in Cameroon. *PLoS One* **2012**; 7:e36954.
  35. Portaels F, Aguiar J, Debacker M, et al. Prophylactic effect of *Mycobacterium bovis* BCG vaccination against osteomyelitis in children with *Mycobacterium ulcerans* disease (Buruli ulcer). *Clin Diagn Lab Immunol* **2002**; 9:1389–91.
  36. BCG vaccination against *Mycobacterium ulcerans* infection (Buruli ulcer): first results of a trial in Uganda. *Lancet* **1969**; 1:111–5.
  37. Trunz BB, Fine P, Dye C. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet* **2006**; 367: 1173–80.



## ARTICLE 5

**Quentin B Vincent**, Marie-Françoise Ardant, Jacques Gnossike, Ambroise Adeye,  
Didier Agossadou, Laurent Marsollier, Annick Chauty, Laurent Abel, Alexandre Alcaïs

**Buruli ulcer of extreme severity maps to the beta-defensin locus on chromosome 8.**

Soumission imminente

Novembre 2014



# **Buruli ulcer of extreme severity maps to the beta-defensin locus on chromosome 8**

Quentin B Vincent<sup>1,2</sup>, Marie-Françoise Ardant<sup>3,4</sup>, Jacques Gnossike<sup>3,4</sup>, Ambroise Adeye<sup>3,4</sup>, Didier Agossadou<sup>5</sup>, Christian Johnson<sup>4</sup>, Laurent Marsollier<sup>6</sup>, Annick Chauty<sup>3,4</sup>, Laurent Abel<sup>1,7</sup>, Alexandre Alcaïs<sup>1,7,8</sup>

(1) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Recherche Médicale U1163 (INSERM),

(2) Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris, France

(3) Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB), BP 191, Pobè, Benin

(4) Fondation Raoul Follereau, Paris, France

(5) Programme de Lutte Contre la Lèpre et l'Ulcère de Buruli, Ministère de la Santé, Cotonou, Bénin

(6) Institut National de la Recherche Médicale U892 (INSERM) et CNRS U6299, équipe 7, Université et CHU d'Angers, Angers, France

(7) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA

(8) CIC-0109 Cochin-Necker Inserm, Unité de Recherche Clinique, Paris Centre Descartes Necker Cochin, Assistance Publique- Hôpitaux de Paris et E.A. 3620, Université Paris Descartes, Paris

**Correspondence** and requests for materials should be addressed to Alexandre Alcaïs or Quentin Vincent, Laboratory of Human Genetics of Infectious Diseases, INSERM U1163, Imagine Institute, 24 bd du Montparnasse, 75015 Paris France

([alexandre.alcais@inserm.fr](mailto:alexandre.alcais@inserm.fr) or [quentin.vincent@inserm.fr](mailto:quentin.vincent@inserm.fr))

Phone :+33 (0) 1 72 60 55 19

## Introduction

Buruli ulcer (BU), caused by *Mycobacterium ulcerans*, is the third most frequent mycobacteriosis worldwide, after tuberculosis and leprosy (1). It mostly affects rural areas of tropical countries. Although no official estimate of global incidence is currently available, West Africa is the principal endemic zone, with a total of more than 4000 new cases reported by Ivory Coast, Ghana and Benin in 2010 (2). BU is a devastating necrotizing skin infection classically characterized by preulcerative lesions (nodules, plaques, oedematous infiltration), eventually developing into deep ulcers with undermined edges. BU causes life-long functional sequelae in more than 20% of patients, most of whom are children (3).

Considerable interindividual variability is observed in the severity of Buruli ulcer. We recently identified severe Buruli ulcer as presentation with oedema, osteomyelitis, or large ( $\geq 15$  cm in diameter), or multifocal lesions (3). Here, we describe an extremely severe form of Buruli ulcer segregating in a large consanguineous family from Benin and report the results from extensive genetic investigation by homozygosity mapping, copy number variant analysis and exome sequencing. These analyses lead to the identification of the beta-defensin locus as a potential key component of the host response to *Mycobacterium ulcerans*.

P1's clinical course is by far the most severe among more than 1,500 BU cases treated over the last 10 years at the CDTUB. The extreme severity of the phenotype together with its inheritance pattern in a consanguineous family strongly suggests recessive Mendelian inheritance with complete penetrance. This hypothesis is further bolstered by the proof-of-concept discoveries of Mendelian forms of other common and rare mycobacterioses, eg, tuberculosis and bccitis (5–7). We investigated this family through genome-wide linkage analysis by homozygosity mapping and copy-number variation analysis. and exome sequencing.

## Subjects and methods

The proband was identified as the most severe Buruli ulcer case among >1,500 patients seen at the Centre de Dépistage et de Traitement de l'Ulcère de Buruli (CDTUB) in Pobè, Benin since its foundation in 2003. Diagnosis was made clinically and confirmed by Ziehl-Nielsen staining and PCR amplification of the specific *IS2404* sequence in multiple samples from multiple lesions, including surgical bone samples. The proband's sister also developed severe PCR-confirmed Buruli ulcer. The parents and seven other sibs were unaffected. The parents were consanguineous (2<sup>nd</sup> degree cousins). The consanguinity in this family strongly suggests a recessive mode of inheritance. We thus performed homozygosity mapping, a model-based method for linkage analysis, which takes into account the consanguinity information. It specifically identifies regions of the genome that are homozygous and identical-by-descent in affected individuals only. Linkage analysis was performed in the MERLIN software version 1.2 using its clustering option to take linkage disequilibrium into account (8,9). A recessive model with complete penetrance was specified.

Genetic research on susceptibility to Buruli ulcer was approved by the institutional review board of the CDTUB (IRB00006860) and the national BU control authorities and informed consent was obtained from all participating family members. Blood was obtained from the 2 unaffected parents, 2 affected and 3 unaffected sibs. DNA was extracted from whole blood following the Nucleon BACC2 Genomic DNA extraction protocol (GE Healthcare), assayed by the QuantIt Picogreen dsDNA kit (Life Technologies) and processed for the genotyping of >900,000 Single Nucleotide Polymorphisms (SNPs) and >900,000 monomorphic nucleotides by the Affymetrix Genome-Wide 6.0 array. Stringent quality control was applied at the individual and SNP levels. Individual quality control included individual call rate (>95%), genetic sex match and genetic familial relationship compatibility. For linkage analysis, within-family SNP quality control included a within-family SNP call rate of 100%, a non-zero



minor allele frequency and the absence of Mendelian errors. As several quality control measures cannot be applied to a single family, we further filtered SNPs on a population basis in the matching Hapmap Yoruba population, and selected SNPs with a call rate of 95% or more, a minor allele frequency of 0.001 or more and Hardy-Weinberg equilibrium at the 0.01 level.

Homozygosity mapping may identify linkage regions in which causal mutations are located, but not the causal mutations themselves. We thus searched for structural variations (insertion and deletions) through Copy-Number Variation (CNV) analysis and for point mutations through exome sequencing of the two patients. CNV identification relies on the quantitative analysis of the intensity of the hybridization signal of nucleotide probes (the logR ratio, LRR) throughout the genome to infer normal, increased or decreased copy-number in the patient's genome at the location of the probe. Genome-wide Copy-Number Variation (CNV) analysis was performed with the PennCNV software version 2011Jun16 (10). To perform exome sequencing, genomic DNA was sheared with a Covaris S2 Ultrasonicator (Covaris). An adaptor-ligated library was prepared with the Paired-End Sample Prep kit V1 (Illumina). Exome capture was performed with the SureSelect Human All Exon v2 kit (Agilent Technologies), covering 38 Mb of the genome. Single-end sequencing was performed on an Illumina Genome Analyzer IIx (Illumina), generating 72 base reads. The sequences were aligned with the human genome reference sequence (hg19 build) with BWA aligner (11). Downstream processing was carried out with the Genome Analysis Toolkit (GATK) (12) SAMtools (13), and Picard Tools (<http://picard.sourceforge.net>). Substitution calls were made with a GATK UnifiedGenotyper, whereas indel calls were made with a GATK IndelGenotyperV2. All calls with a read coverage  $\leq 2\times$  and a Phred-scaled SNP quality of  $\leq 20$  were filtered out. All the variants were annotated with the GATK Genomic Annotator.

## Results

### *Case report*

The index case patient P1 was born in 2000 from consanguineous parents (**Figure 1**). At 5 years of age, she presented at the CDTUB in Pobè, Benin, with cachexia, fever and an oedematous lower right limb (**Figure 2**). *M. ulcerans* was identified in the synovial fluid of the right knee by Ziehl-Nielsen staining and PCR amplification. She also presented with PCR-positive oedema of the left foot. Osteo-arthritis of the right tibia and knee and osteomyelitis of the left cuboid were confirmed by X-ray. Despite prolonged treatment by rifampicin and streptomycin (90 days) and multiple surgeries, the infection further disseminated. At 2 months after diagnosis, oedema of the right arm revealed right elbow osteo-arthritis, further confirmed by the aspiration of caseiform matter, Ziehl-Nielsen staining and PCR amplification. Fever did not recede before the amputation of her right leg at 3 months. At 6 months, osteomyelitis of the left radius was diagnosed. All four limbs were thus involved at this point. At one year, the left fibula was involved. She underwent numerous iterative surgeries of all limbs in the first 18 months after diagnosis. She relapsed two, three and five years after the initial diagnosis, with respective involvement of the right humerus, left calcaneal tendon and left tibia. She had no infectious relapse in the past 4 years. *Mycobacterium ulcerans* was repeatedly identified by Ziehl-Nielsen staining and *IS2404* PCR on different lesions at different times on all four limbs, including surgical bone samples (**Figure 2**). The medical history of P1 also includes several malaria episodes and HBV hepatitis. P1's sister, P2, was born in 1995. At 13 years of age, she presented with a right oedematous arm. PCR and culture were positive for *Mycobacterium ulcerans*. Both patients were HIV-negative and had been vaccinated against BCG at birth. Their complete blood count was normal. The parents and 7 other sibs were unaffected.

### *Genetic investigations*

After quality checks, 310993 SNPs remained and were grouped in 126704 independent clusters for model-based linkage analysis by homozygosity mapping, reaching a genome-wide mean information content of 0.72. A theoretical maximum LOD score can be calculated given a familial configuration, a completely informative marker and a genetic model. In our case (2 affected sibs, 3 unaffected sibs, consanguinity loop at 2<sup>nd</sup> degree cousins and a recessive model with complete penetrance), this maximum LOD score is of 2.8, translating to a minimal p-value of  $3.10^{-3}$ . Seven regions reached the theoretical maximum LOD score and mapped to chromosomes 2, 5, 7 and 8 (**Figure 3**), spanning a total of 5.4 megabases. The chromosome 2 region and the second region on chromosome 7 in fact approached the maximal LOD score, with respective LOD scores at 2.3 and 2.5, to be pondered by their information content of 0.73 and 0.75. The other regions all had a LOD score above 2.7 and information content above 0.85. The complete list of genes in each of these regions (retrieved on Ensembl v75) is available in the **Supplemental Table**. Interestingly, two of the three linkage regions on chromosome 8 contain clusters of genes coding for beta-defensins.

Based on a recessive mode of inheritance, we identified 5,397 exonic homozygous variations shared by both patients. After filtering out known variations from public databases, 9 remained (Supplementary Table). None were located in the linkage regions and none appeared to be interesting candidates. Mapping quality was low in the defensin region and precluded variation calling.

Whole-genome copy-number variation analysis revealed an 8.5kb homozygous deletion shared by both patients. The deletion segregates perfectly with the phenotype, as it is heterozygous in both parents, heterozygous in one unaffected sib, and absent in two

unaffected sibs (**Figure 4**). No other CNV displayed these characteristics genome-wide. Additionally, the deletion is located in the first linked region identified on chromosome 8.

## Discussion

We report the clinical phenotype and genetic analysis of a consanguineous family with Buruli ulcer of extreme severity displaying Mendelian inheritance. The index patient suffered from aggressive multifocal *Mycobacterium ulcerans* osteoarthritis that resisted antibiotic and surgical treatment over a period of more than 2 years, relapsed several times over 5 years and resulted in severe handicap including amputation of the right leg. Her sister suffered from severe oedematous Buruli ulcer of the right arm. We identified a homozygous deletion in a beta-defensin locus on chromosome 8 in both patients.

Beta-defensins are a family of antimicrobial peptides of innate immunity, which are widely expressed throughout epithelial tissues in response to infectious agents (14,15). Interestingly, beta-defensins have also been implicated in the normal healing process of an aseptic skin wound (15,16). These antibacterial and wound repair functions of beta-defensins support their potential role in the human response to *Mycobacterium ulcerans*. They have been shown *in vitro* and *in vivo* to be upregulated in response to several mycobacteria (17–22), including *M. ulcerans* (23) and have specifically been implicated in the immune response to bone infection both in mice and humans (24–26).

Their genetic organization is of great complexity, with many paralogs found in the human genome (27). Ensembl v75 reports 38 beta-defensin genes on 4 chromosomes (mainly chromosome 8 and 20) and 11 pseudogenes. The beta-defensin locus on chromosome 8p23 comprising beta-defensin genes *DEFB4*, *DEFB103*, *DEFB104*, *DEFB105*, *DEFB106*, *DEFB107* has wide copy-number variability in the human population, with most individuals

carrying four copies of this cluster (28). Interestingly, no individual was ever reported to have zero copy of the cluster, highlighting its capital role in the human immune response. We report linkage with this repeat as linkage region 1 on chromosome 8. We also report a ~10kb deletion within linkage region 2 on chromosome 8. This second linkage region (~4.5 Mb downstream from the first linkage region) contains defensins *DEFB130*, *DEFB134*, *DEFB135* and *DEFB136*, and is not known to be copy-number variable in the human population.

CNV detection algorithms have a modest resolution at determining CNV breakpoints and are known to underestimate CNV size, which means that the size of the identified deletion could be larger than 10kb (29). Interestingly, two microRNAs have been identified less than 40 kb from the deletion. Earlier versions of Ensembl described a long non-coding RNA at the exact location of the deletion, which is however absent from current version v75. The regulatory potential of non-coding RNAs might well mean that beta-defensin expression is abnormal in the patients we investigated. As a final step in the genetic investigation of these patients, we performed exome sequencing in both patients, which did not identify homozygous mutations shared by the two sisters in the linkage regions. The existence of such mutations is not entirely excluded in the beta-defensin linkage regions, as we faced severe alignment issues in the beta-defensin regions. Extreme sequence repetition indeed precluded genotype calling by the dedicated algorithms (30). This highlights one of the important limits of exome sequencing.

This unique clinical case highlights the critical interindividual variability in the human response to *Mycobacterium ulcerans*. In highly endemic foci of infection, most exposed individuals do not develop Buruli ulcer lesions, whereas others, like P2, rapidly develop extensive skin lesions, and others, like P1, rapidly develop multifocal skin and bone lesions despite aggressive medical and surgical treatment. The origin of this interindividual variability might be environmental but also genetic, an hypothesis supported by the

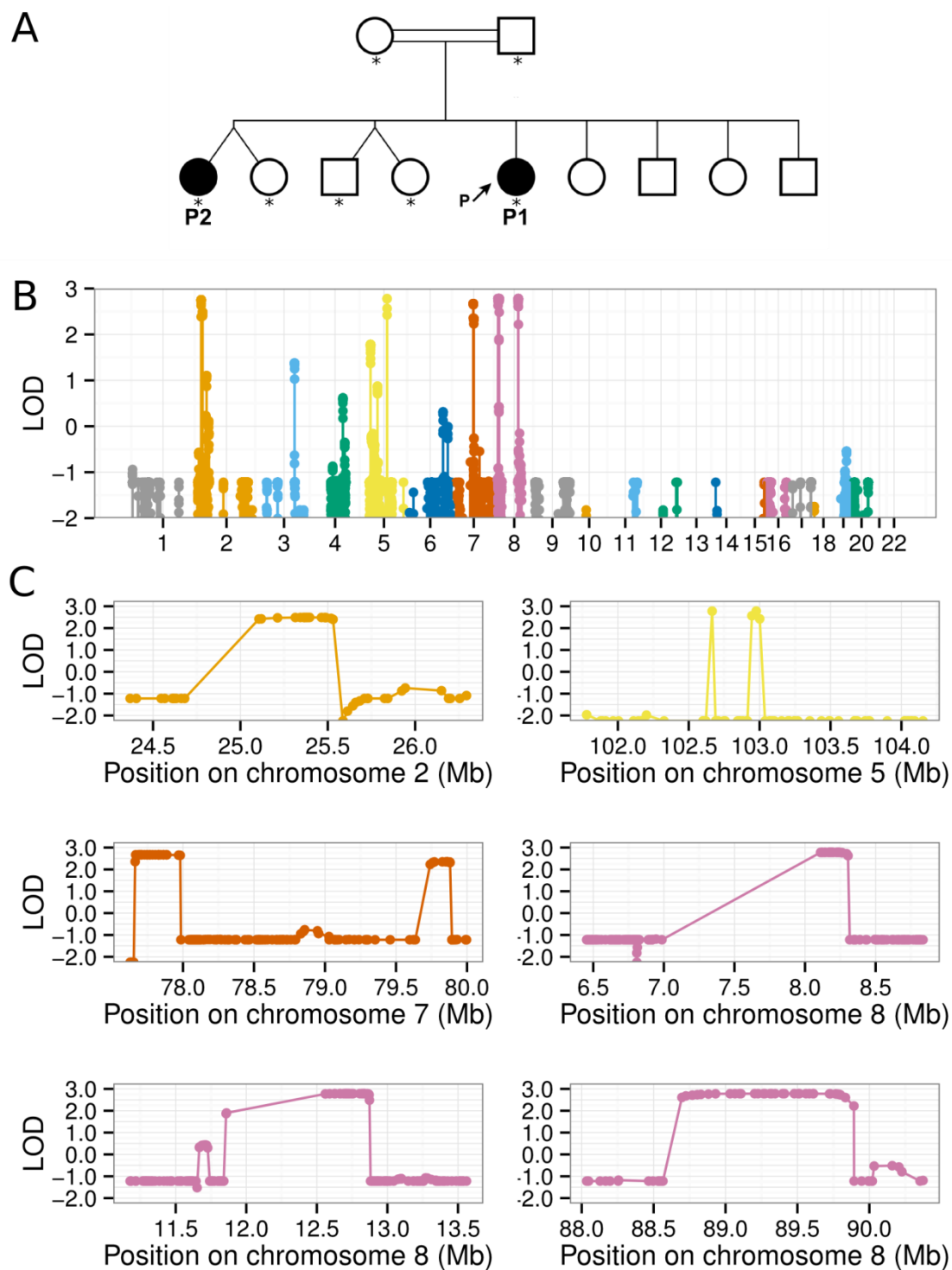
identification of a linkage region containing beta-defensins in this large multiplex family from Benin. Further functional investigation of this family is required to definitively confirm this hypothesis but needs to face the non-trivial challenges posed by the complexity of this dynamic region of the human genome.

**Figure 1. Clinical evolution of a case of BU of extreme severity (P1).**

YEAR	AGE	CLINICAL EVENTS
2000		<b>Birth</b>
2005	5 y.o	<b>Initial Diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: right thigh and leg; left foot</li> <li>- Septic Arthritis: right knee</li> <li>- Osteomyelitis : right fibula and tibia; left cuboid</li> <li>- Lab: Ziehl-Neelsen and PCR+ on both limbs</li> </ul> <b>Dissemination at 2 months post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: right arm and forearm</li> <li>- Septic Arthritis: right elbow</li> <li>- Osteomyelitis : right humerus and radius and ulna</li> <li>- Lab: Ziehl-Neelsen and PCR +</li> </ul> <div style="border: 1px solid red; padding: 5px; margin: 10px 0;"> <b>3 months post-diagnosis</b>  Amputation of the lower right limb </div> <b>Dissemination at 6 months post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: left forearm</li> <li>- Osteomyelitis : left radius</li> <li>- Lab: Ziehl-Neelsen +; PCR not performed</li> </ul>
2006	6 y.o	<b>Dissemination at 1 year post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: left leg</li> <li>- Osteomyelitis : left fibula</li> <li>- Lab: Ziehl-Neelsen and PCR +</li> </ul>
2007	7 y.o	<b>Dissemination at 2 years post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: right arm</li> <li>- Osteomyelitis : right humerus</li> <li>- Lab: Ziehl-Neelsen +; PCR not performed</li> </ul>
2008	8 y.o	<b>Dissemination at 3 years post-diagnosis</b> <ul style="list-style-type: none"> <li>- Soft tissue collection: left calcaneal tendon</li> </ul>
2010	10 y.o	<b>Dissemination at 5 years post-diagnosis</b> <ul style="list-style-type: none"> <li>- Oedema: left leg</li> <li>- Osteomyelitis : left tibia</li> <li>- Lab: Ziehl-Neelsen and PCR +</li> </ul> <b>No further clinical event since 2010</b>

**Figure 2. BU of extreme severity segregates in a consanguineous family from Benin and maps to chromosomes 5 and 8.**

- A. Pedigree tree. Genotyped individuals are indicated by stars.
- B. Genome-Wide Linkage Analysis by Homozygosity Mapping
- C. Details of the regions of interest on chromosomes 2, 5, 7 and 8.

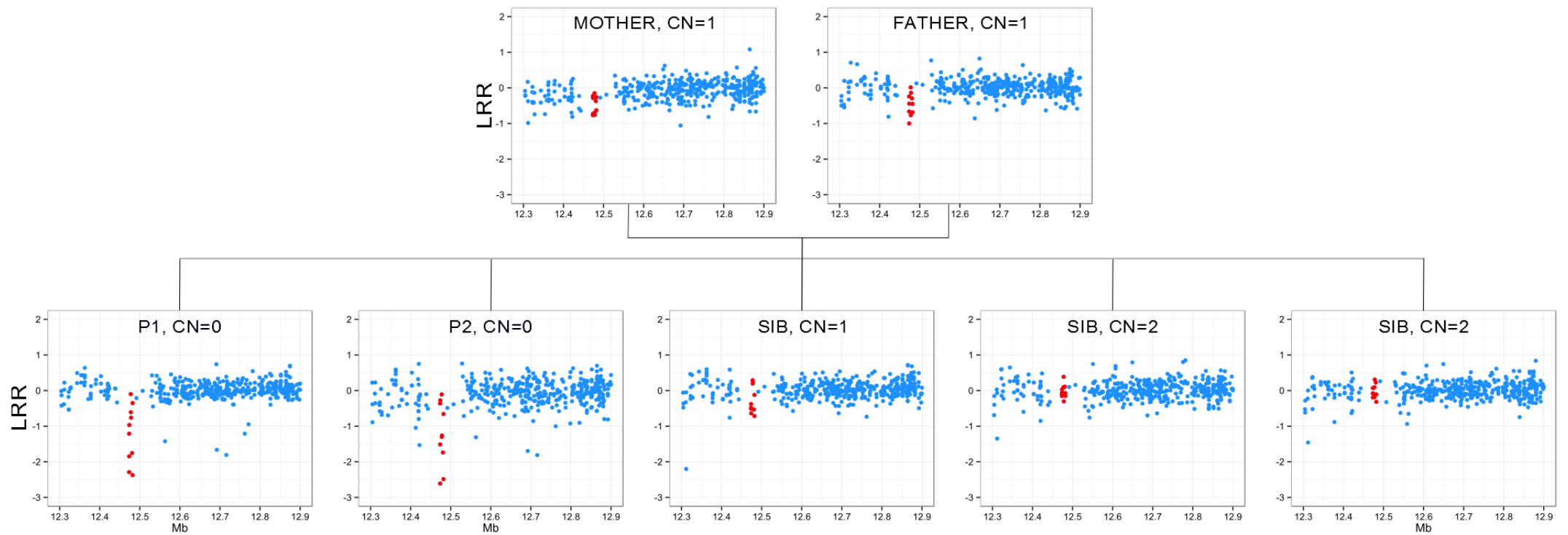




**Figure 3. Genome-Wide CNV analysis identifies a homozygous deletion in the second linkage region on chromosome 8.**

LRR: logR-ratio, a normalized quantification of the binding intensity of array probes.

CN: Estimated Copy Number of the chromosomal segment (shown in red).



**Supplemental Table. List of genes within linkage regions (1/2)**

<b>Chr2</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>NCOA1</b>	nuclear receptor coactivator 1	protein_coding
<b>PTRHD1</b>	peptidyl-tRNA hydrolase domain containing 1	protein_coding
<b>CENPO</b>	centromere protein O	protein_coding
<b>ADCY3</b>	adenylate cyclase 3	protein_coding
<b>DNAJC27</b>	DnaJ (Hsp40) homolog, subfamily C, member 27	protein_coding
<b>EFR3B</b>	EFR3 homolog B ( <i>S. cerevisiae</i> )	protein_coding
<b>RP11-509E16.1</b>		lincRNA
<b>POMC</b>	proopiomelanocortin	protein_coding
<b>RP11-458N5.1</b>		lincRNA
<b>DNMT3A</b>	DNA (cytosine-5-)-methyltransferase 3 alpha	protein_coding
<b>MIR1301</b>	microRNA 1301	miRNA

<b>Chr7</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>MAGI2</b>	membrane associated guanylate kinase	protein_coding
<b>GNAI1</b>	guanine nucleotide binding protein (G protein), alpha inhibiting	protein_coding
<b>AC003988.1</b>		lincRNA

<b>Chr8.1</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>AF228730.2</b>		miRNA
<b>FAM66B</b>	family with sequence similarity 66, member B	lincRNA
<b>ZNF705G</b>	zinc finger protein 705G	protein_coding
<b>DEFB4B</b>	defensin, beta 4B	protein_coding
<b>DEFB103B</b>	defensin, beta 103B	protein_coding
<b>SPAG11B</b>	sperm associated antigen 11B	protein_coding
<b>AC130360.1</b>		miRNA
<b>DEFB104B</b>	defensin, beta 104B	protein_coding
<b>DEFB106B</b>	defensin, beta 106B	protein_coding
<b>DEFB105B</b>	defensin, beta 105B	protein_coding
<b>DEFB107B</b>	defensin, beta 107B	protein_coding
<b>PRR23D1</b>	proline rich 23 domain containing 1	protein_coding
<b>PRR23D2</b>	proline rich 23 domain containing 2	protein_coding
<b>DEFB107A</b>	defensin, beta 107A	protein_coding
<b>DEFB105A</b>	defensin, beta 105A	protein_coding
<b>DEFB106A</b>	defensin, beta 106A	protein_coding
<b>DEFB104A</b>	defensin, beta 104A	protein_coding
<b>SPAG11A</b>	sperm associated antigen 11A	protein_coding
<b>AC130365.1</b>		miRNA
<b>DEFB103A</b>	defensin, beta 103A	protein_coding
<b>DEFB4A</b>	defensin, beta 4A	protein_coding
<b>ZNF705B</b>	zinc finger protein 705B	protein_coding
<b>FAM66E</b>	family with sequence similarity 66, member E	lincRNA
<b>MIR548I3</b>	microRNA 548i-3	miRNA

**Supplemental Table. List of genes within linkage regions (2/2)**

<b>Chr8.1 (...)</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>AC068020.1</b>		miRNA
<b>LRLE1</b>	Liver-related low express protein 1	protein_coding
<b>SGK223</b>	Tyrosine-protein kinase Sgk223	protein_coding
<b>CTA-398F10.1</b>		lincRNA
<b>CTA-398F10.2</b>		lincRNA

<b>Chr8.2</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>DEFB135</b>	defensin, beta 135	protein_coding
<b>DEFB134</b>	defensin, beta 134	protein_coding
<b>RP11-481A20.11</b>	Protein LOC101060662	protein_coding
<b>DEFB130</b>	beta-defensin 130-like precursor	protein_coding
<b>ZNF705D</b>	zinc finger protein 705D	protein_coding
<b>USP17L2</b>	ubiquitin specific peptidase 17-like family member 2	protein_coding
<b>AC145124.1</b>		miRNA
<b>FAM86B1</b>	family with sequence similarity 86, member B1	protein_coding
<b>DEFB130</b>	defensin, beta 130	protein_coding
<b>LINC00965</b>	long intergenic non-protein coding RNA 965	lincRNA
<b>AC087203.1</b>		miRNA
<b>FAM86B2</b>	family with sequence similarity 86, member B2	protein_coding
<b>RP11-419I17.1</b>		lincRNA
<b>AC130352.1</b>		miRNA
<b>AC068587.2</b>		miRNA
<b>AC068587.1</b>		miRNA
<b>MIR5692A1</b>	microRNA 5692a-1	miRNA
<b>LONRF1</b>	LON peptidase N-terminal domain and ring finger 1	protein_coding
<b>MIR3926-2</b>	microRNA 3926-2	miRNA
<b>RP11-252C15.1</b>		lincRNA
<b>LINC00681</b>	long intergenic non-protein coding RNA 681	lincRNA
<b>AC090133.1</b>		miRNA
<b>KIAA1456</b>	KIAA1456	protein_coding

<b>Chr8.3</b>	<b>Description</b>	<b>Gene Biotype</b>
<b>CNBD1</b>	cyclic nucleotide binding domain containing 1	protein_coding
<b>DCAF4L2</b>	DDB1 and CUL4 associated factor 4-like 2	protein_coding
<b>MMP16</b>	matrix metalloproteinase 16 (membrane-inserted)	protein_coding
<b>RP11-642C5.1</b>		lincRNA

Genes coding for a protein, a non-coding RNA of type miRNA or lincRNA were listed here. No gene was retrieved from chromosome 5. The exact intervals used for query of Ensembl75 (GRCh37) are : 2:24680948-25585907, 5:102620674-102689268, 5:102913885-103038495, 7:77653409-77985583, 7:79635822-79894010, 8:6986241-8316581, 8:11840071-12880628, 8:88564546-89894046.

## References

1. Sizaire V, Nackers F, Comte E, Portaels F. *Mycobacterium ulcerans* infection: control, diagnosis, and treatment. *Lancet Infect Dis.* mai 2006;6(5):288-296.
2. WHO | Buruli ulcer disease factsheet [Internet]. WHO. [cité 3 avr 2012]. Disponible sur: <http://www.who.int/mediacentre/factsheets/fs199/en/index.html>
3. Vincent QB, Ardant M-F, Adeye A, Goundote A, Saint-André J-P, Cottin J, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health.* juill 2014;2(7):e422-e430.
4. Debacker M, Aguiar J, Steunou C, Zinsou C, Meyers WM, Guédénon A, et al. *Mycobacterium ulcerans* disease (Buruli ulcer) in rural hospital, Southern Benin, 1997-2001. *Emerg Infect Dis.* août 2004;10(8):1391-1398.
5. Alcaïs A, Fieschi C, Abel L, Casanova J-L. Tuberculosis in children and adults: two distinct genetic diseases. *J Exp Med.* 19 déc 2005;202(12):1617-1621.
6. Casanova J-L, Abel L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu Rev Genomics Hum Genet.* 2013;14:215-243.
7. Bustamante J, Arias AA, Vogt G, Picard C, Galicia LB, Prando C, et al. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. *Nat Immunol.* 2011;12(3):213-21.
8. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002;30(1):97-101.
9. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science.* 1987;236(4808):1567-70.
10. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* nov 2007;17(11):1665-1674.
11. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl.* 15 juill 2009;25(14):1754-1760.
12. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* sept 2010;20(9):1297-1303.
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl.* 15 août 2009;25(16):2078-2079.

14. Lai Y, Gallo RL. AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends Immunol.* mars 2009;30(3):131-141.
15. Semple F, Dorin JR.  $\beta$ -Defensins: Multifunctional Modulators of Infection, Inflammation and More? *J Innate Immun.* 2012;4(4):337-348.
16. Ahrens K, Schunck M, Podda G-F, Meingassner J, Stuetz A, Schröder J-M, et al. Mechanical and metabolic injury to the skin barrier leads to increased expression of murine  $\beta$ -defensin-1, -3, and -14. *J Invest Dermatol.* févr 2011;131(2):443-452.
17. Méndez-Samperio P. Role of antimicrobial peptides in host defense against mycobacterial infections. *Peptides.* oct 2008;29(10):1836-1841.
18. Kisich KO, Heifets L, Higgins M, Diamond G. Antimycobacterial agent based on mRNA encoding human beta-defensin 2 enables primary macrophages to restrict growth of *Mycobacterium tuberculosis*. *Infect Immun.* avr 2001;69(4):2692-2699.
19. Rivas-Santiago B, Sada E, Tsutsumi V, Aguilar-Leon D, Contreras JL, Hernandez-Pando R. beta-Defensin gene expression during the course of experimental tuberculosis infection. *J Infect Dis.* 1 sept 2006;194(5):697-701.
20. Cervantes-Villagrana AR, Hernández-Pando R, Biragyn A, Castañeda-Delgado J, Bodogai M, Martínez-Fierro M, et al. Prime-boost BCG vaccination with DNA vaccines based in  $\beta$ -defensin-2 and mycobacterial antigens ESAT6 or Ag85B improve protection in a tuberculosis experimental model. *Vaccine.* 11 janv 2013;31(4):676-684.
21. Klug-Micu GM, Stenger S, Sommer A, Liu PT, Krutzik SR, Modlin RL, et al. CD40 ligand and interferon- $\gamma$  induce an antimicrobial response against *Mycobacterium tuberculosis* in human monocytes. *Immunology.* mai 2013;139(1):121-128.
22. Cogen AL, Walker SL, Roberts CH, Hagge DA, Neupane KD, Khadge S, et al. Human beta-defensin 3 is up-regulated in cutaneous leprosy type 1 reactions. *PLoS Negl Trop Dis.* 2012;6(11):e1869.
23. Ortiz RH, Leon DA, Estevez HO, Martin A, Herrera JL, Romo LF, et al. Differences in virulence and immune response induced in a murine model by isolates of *Mycobacterium ulcerans* from different geographic areas. *Clin Exp Immunol.* août 2009;157(2):271-281.
24. Zhu C, Wang J, Cheng T, Li Q, Shen H, Qin H, et al. The potential role of increasing the release of mouse  $\beta$ -defensin-14 in the treatment of osteomyelitis in mice: a primary study. *PloS One.* 2014;9(1):e86874.
25. Zhu C, Qin H, Cheng T, Tan H-L, Guo Y-Y, Shi S-F, et al. *Staphylococcus aureus* supernatant induces the release of mouse  $\beta$ -defensin-14 from osteoblasts via the p38 MAPK and NF- $\kappa$ B pathways. *Int J Mol Med.* juin 2013;31(6):1484-1494.
26. Varoga D, Tohidnezhad M, Paulsen F, Wruck CJ, Brandenburg L, Mentlein R, et al. The role of human beta-defensin-2 in bone. *J Anat.* déc 2008;213(6):749-757.

27. Hollox EJ, Armour JAL, Barber JCK. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am J Hum Genet.* sept 2003;73(3):591-600.
28. Hardwick RJ, Machado LR, Zuccherato LW, Antolinos S, Xue Y, Shawa N, et al. A worldwide analysis of beta-defensin copy number variation suggests recent selection of a high-expressing DEFB103 gene copy in East Asia. *Hum Mutat* [Internet]. 8 mars 2011 [cité 19 avr 2011]; Disponible sur: <http://www.ncbi.nlm.nih.gov/pubmed/21387465>
29. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol.* juin 2011;29(6):512-520.
30. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet.* janv 2012;13(1):36-46.



# GENETIQUE COMPLEXE DE L'ULCERE DE BURULI :

## DETAIL DES MANHATTAN PLOTS

### SNPs GENOTYPES

**Sont successivement représentés les Manhattan plots suivants :**

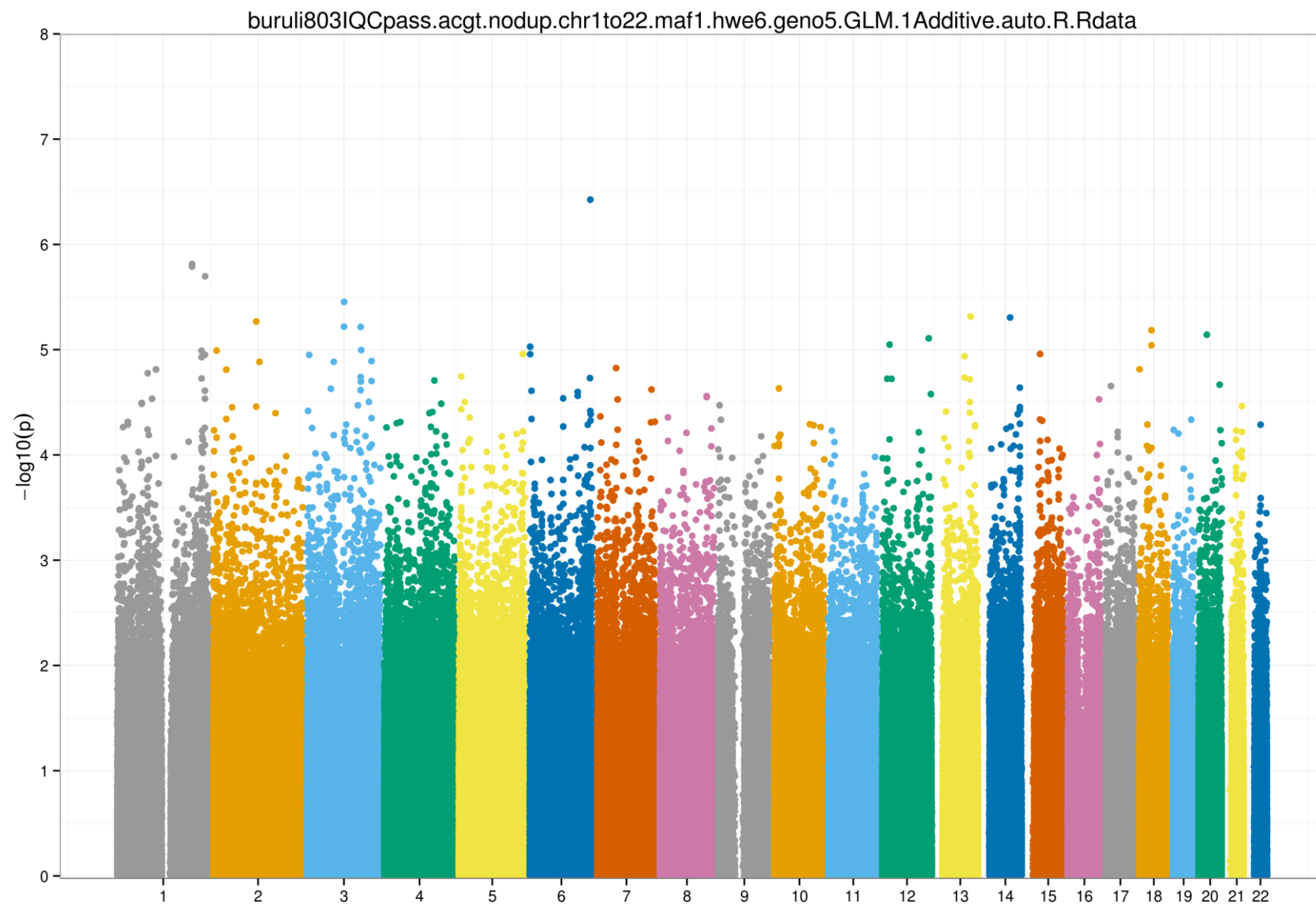
- Modèle linéaire généralisé de type régression logistique (GLM)
- Modèle linéaire généralisé de type régression logistique avec ajustement sur les quatre premières composantes principales (GLMpc1to4)
- Modèle mixte incorporant une matrice de similarité génétique (GEMMA)
- Analyse de survie selon un modèle de Cox (Cox)
- Analyse de survie selon un modèle de Cox incorporant une matrice de similarité génétique (CoxME)

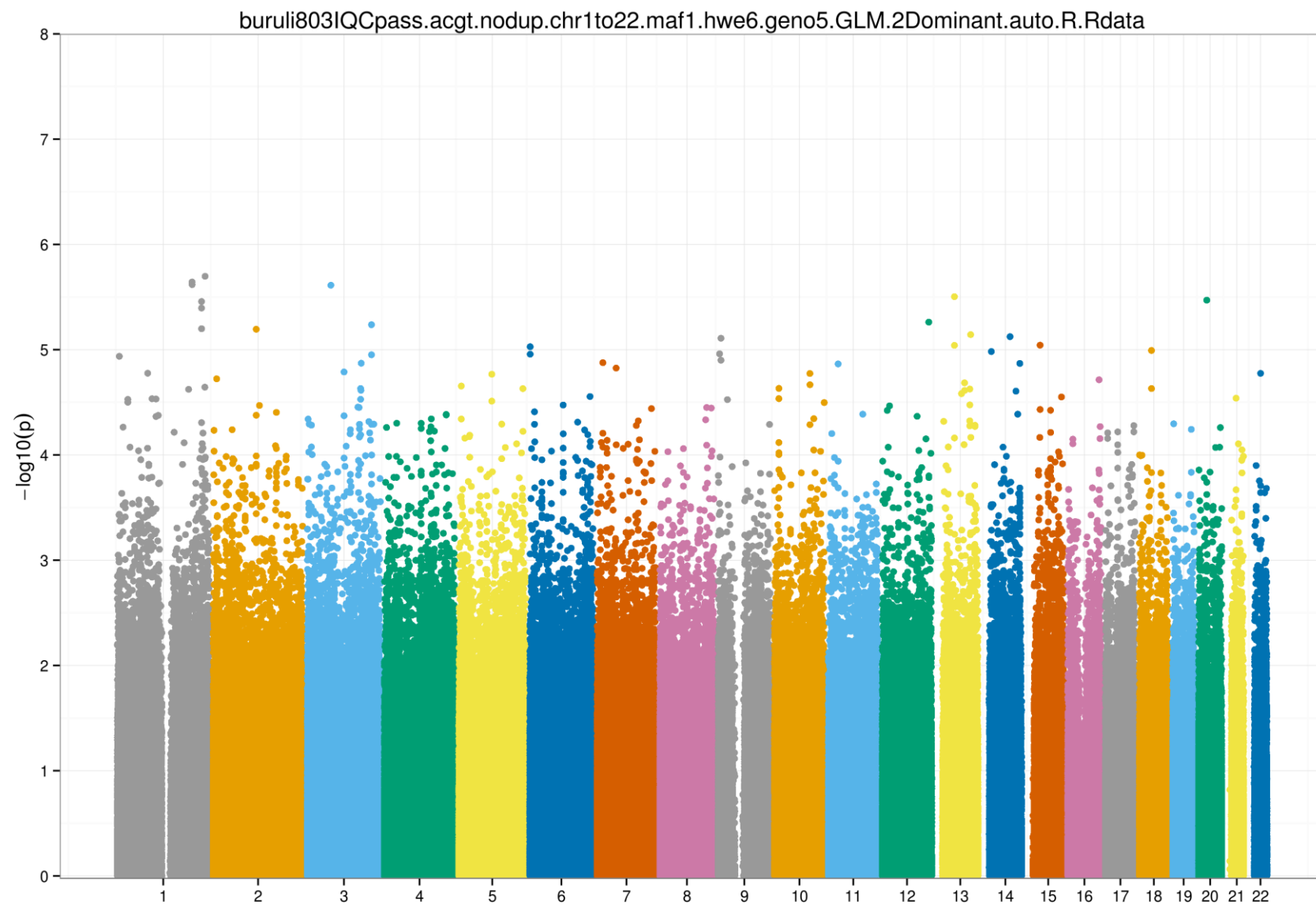
Pour chaque modèle statistique, sont successivement représentés les modèles génétiques additifs, dominants, récessifs, génotypiques (à 2 degrés de liberté), avantage aux hétérozygotes. Le modèle génotypique n'a pas été exploré en modèle mixte.

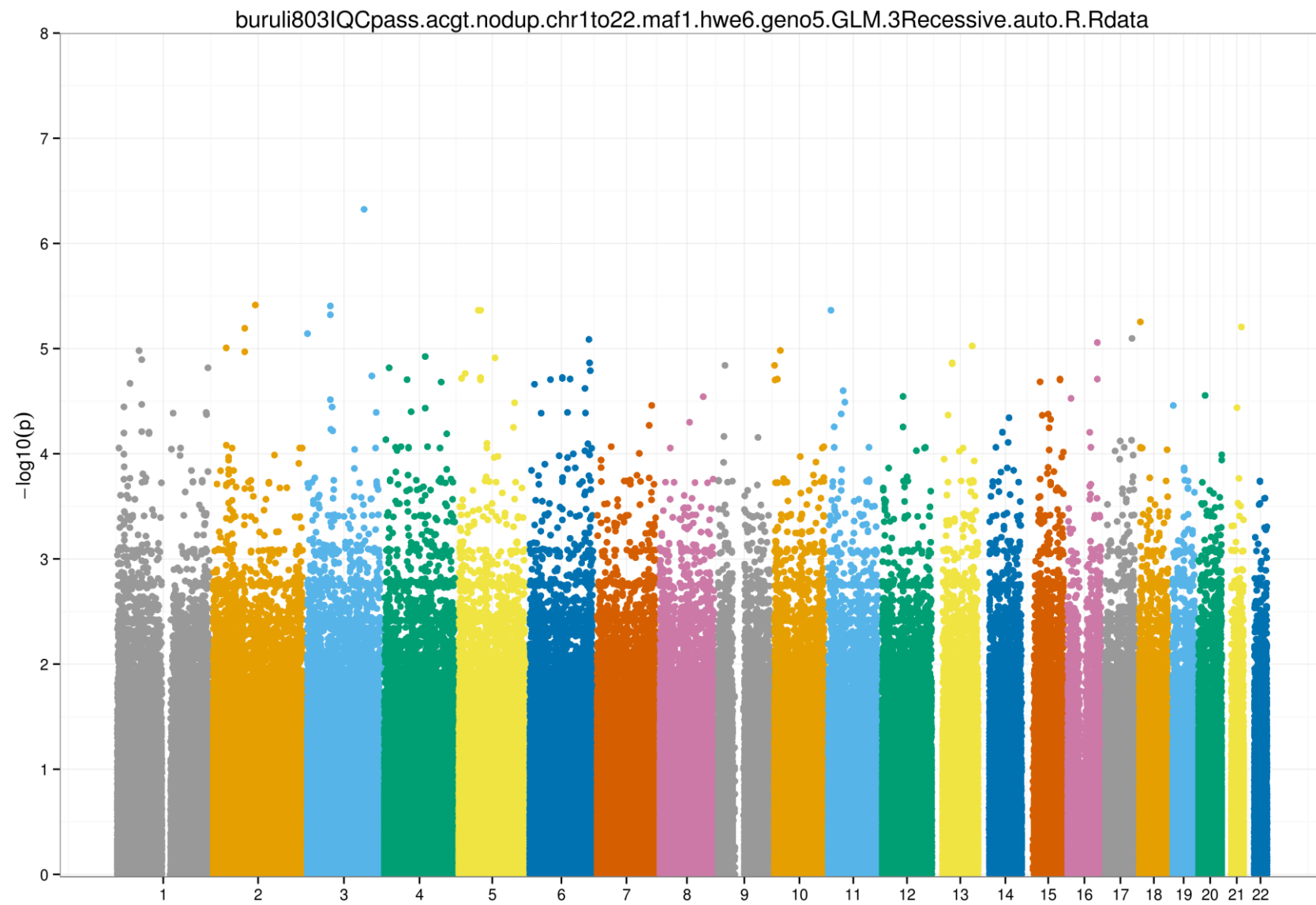
Tous les Manhattan plots représentent l'ensemble des SNPs avec un taux de génotypage supérieur à 95%, une fréquence de l'allèle mineur supérieure à 1%, un respect de l'équilibre de Hardy-Weinberg avec une p-value supérieure à  $10^{-6}$ , soit 1,802,157 marqueurs issus de la puce Illumina Omni2.5-8.

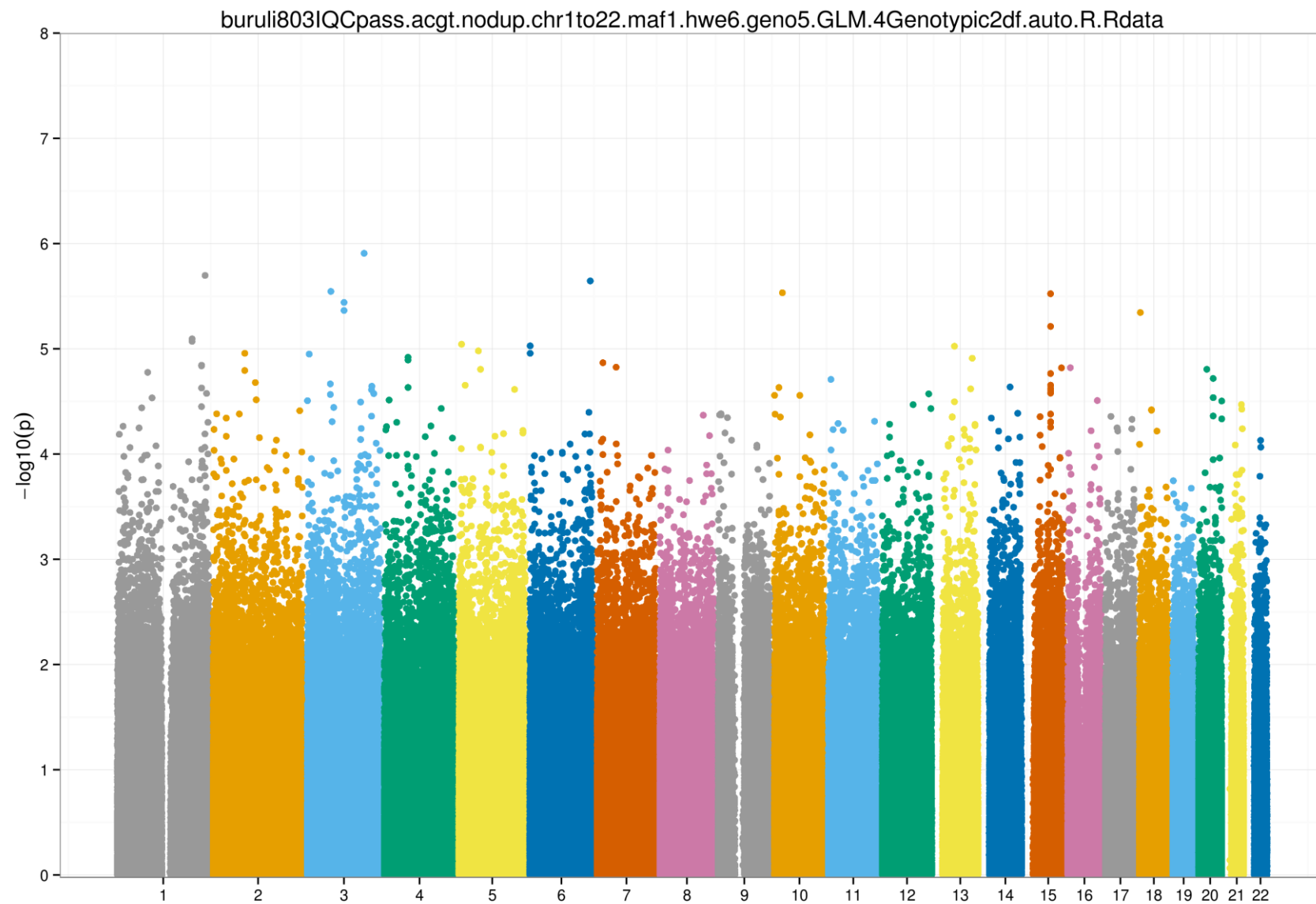


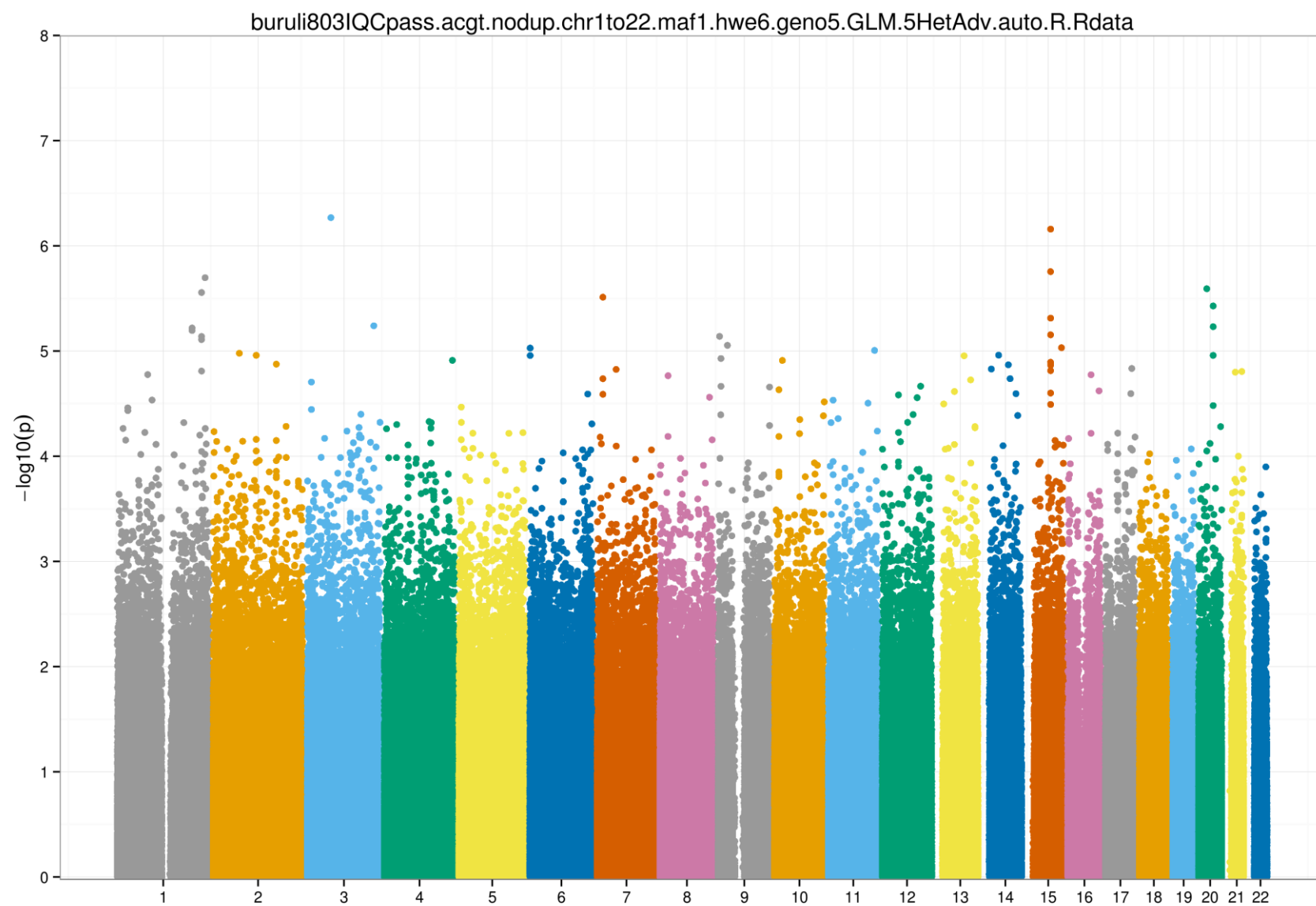


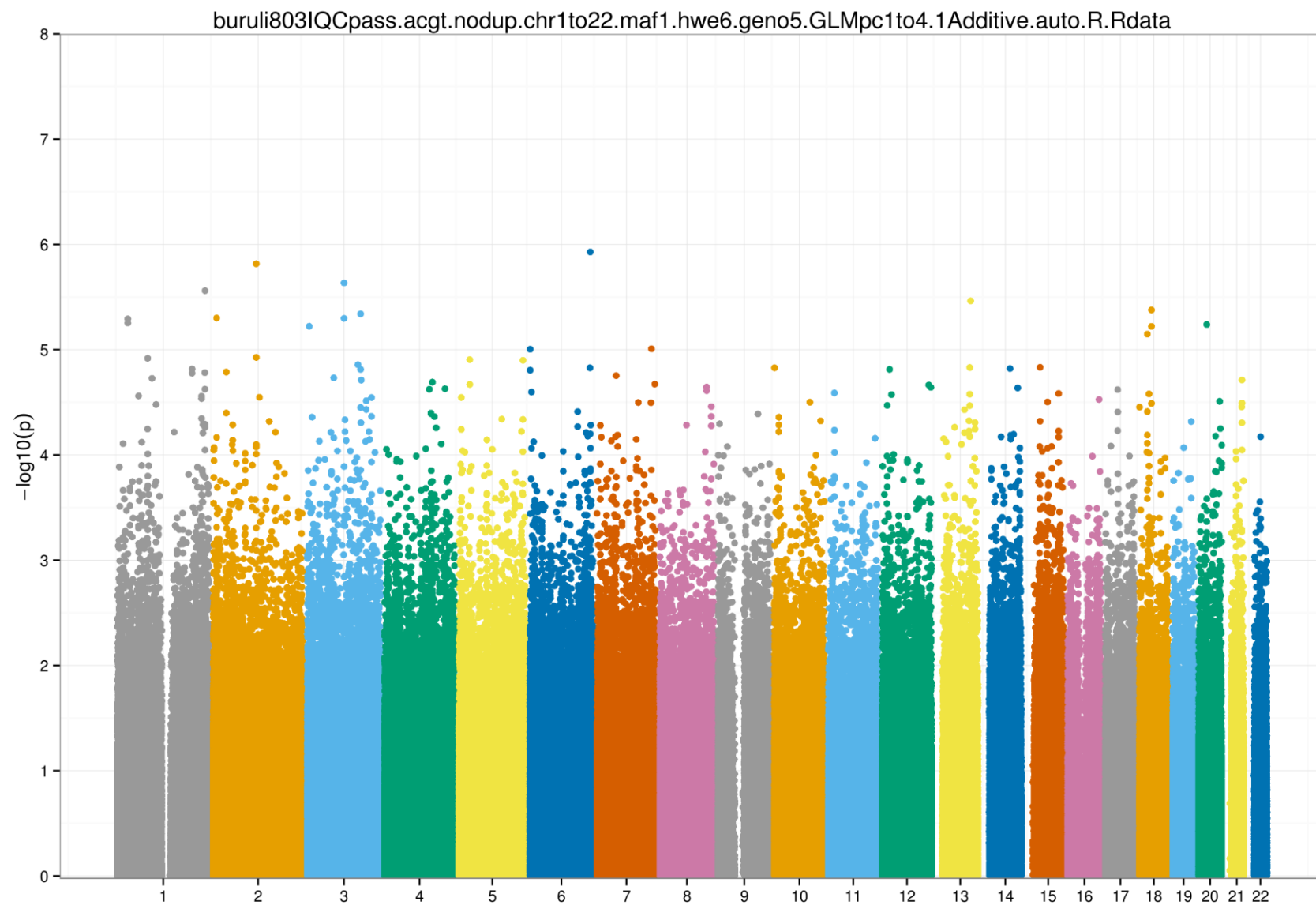


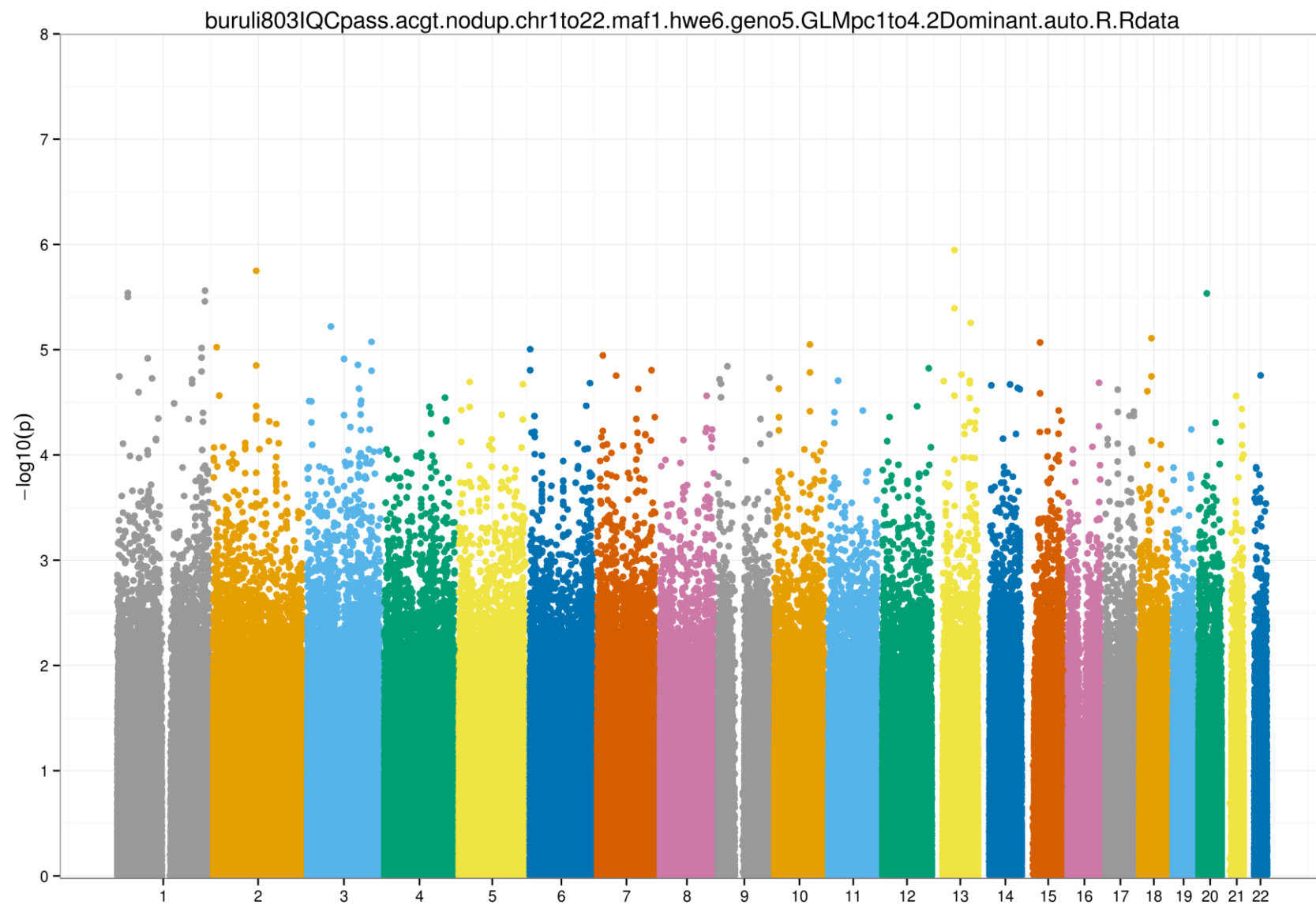




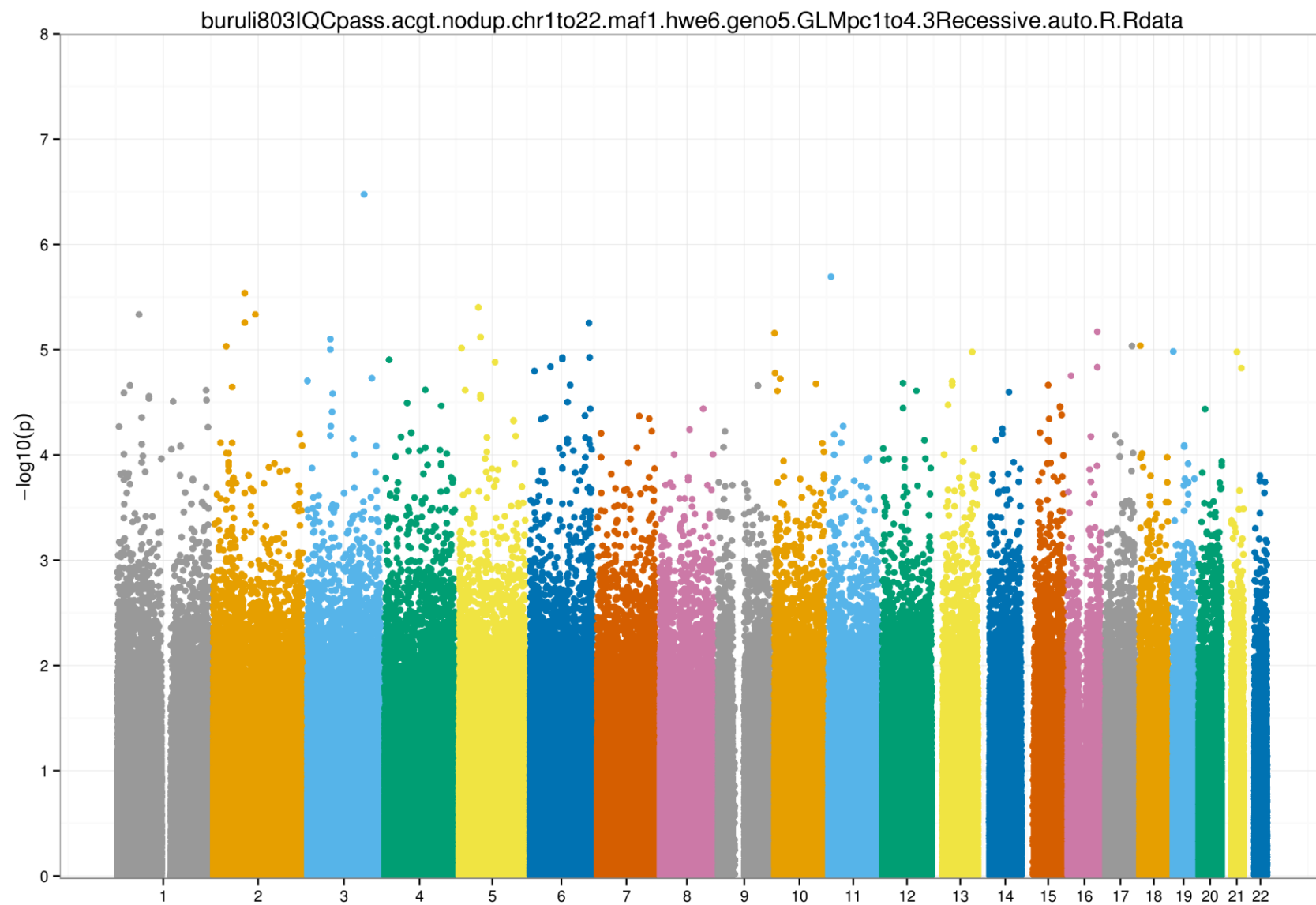


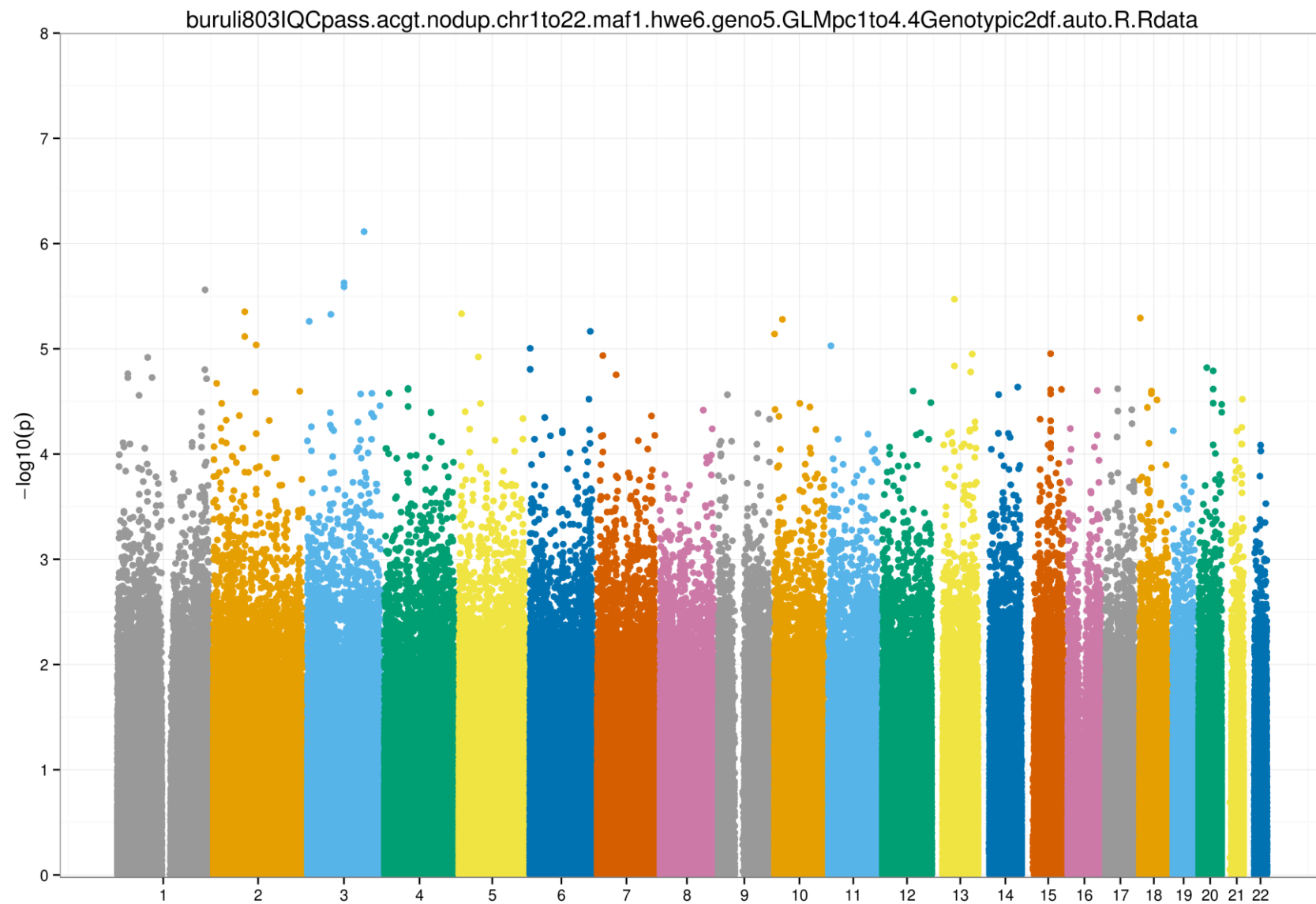


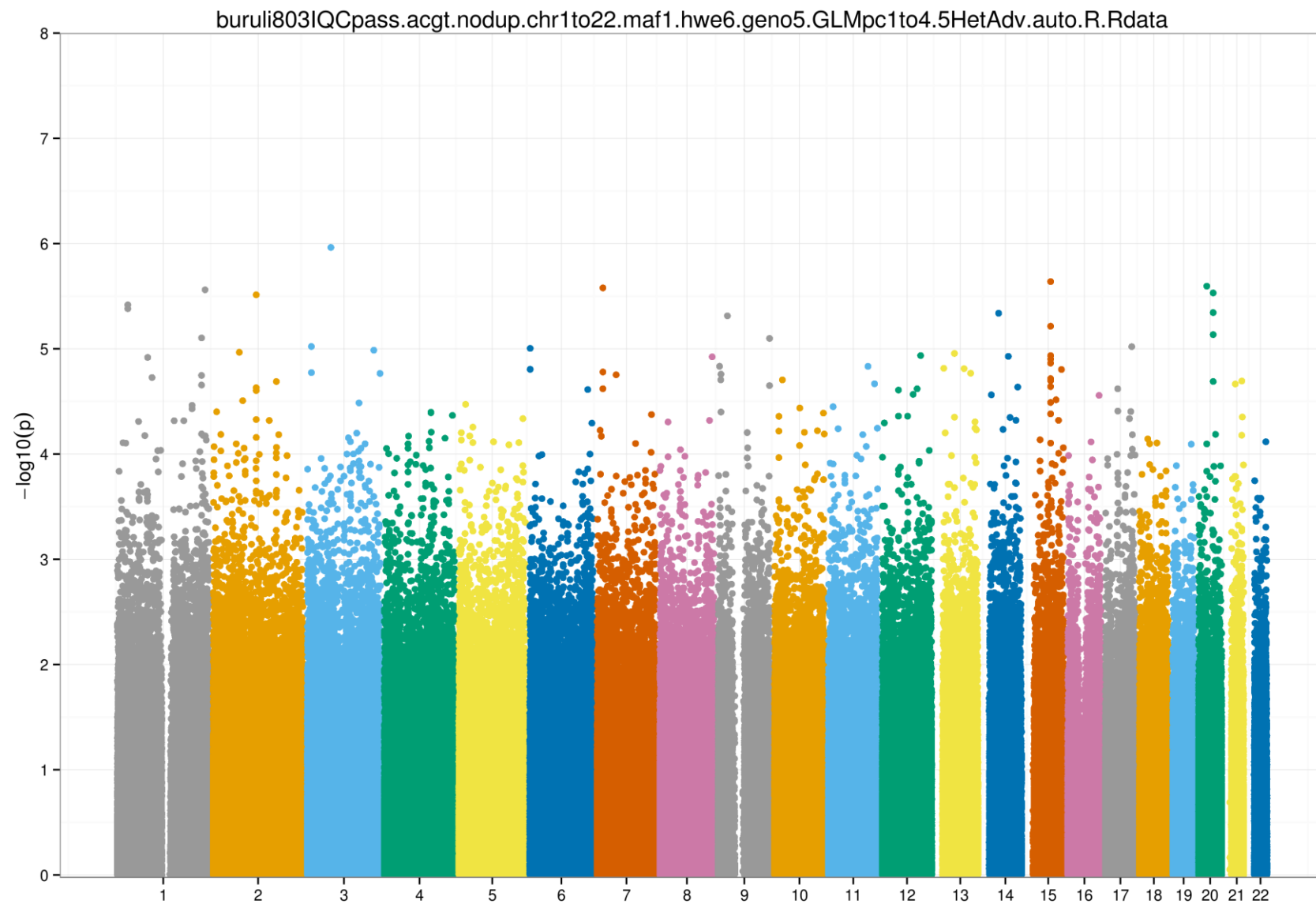


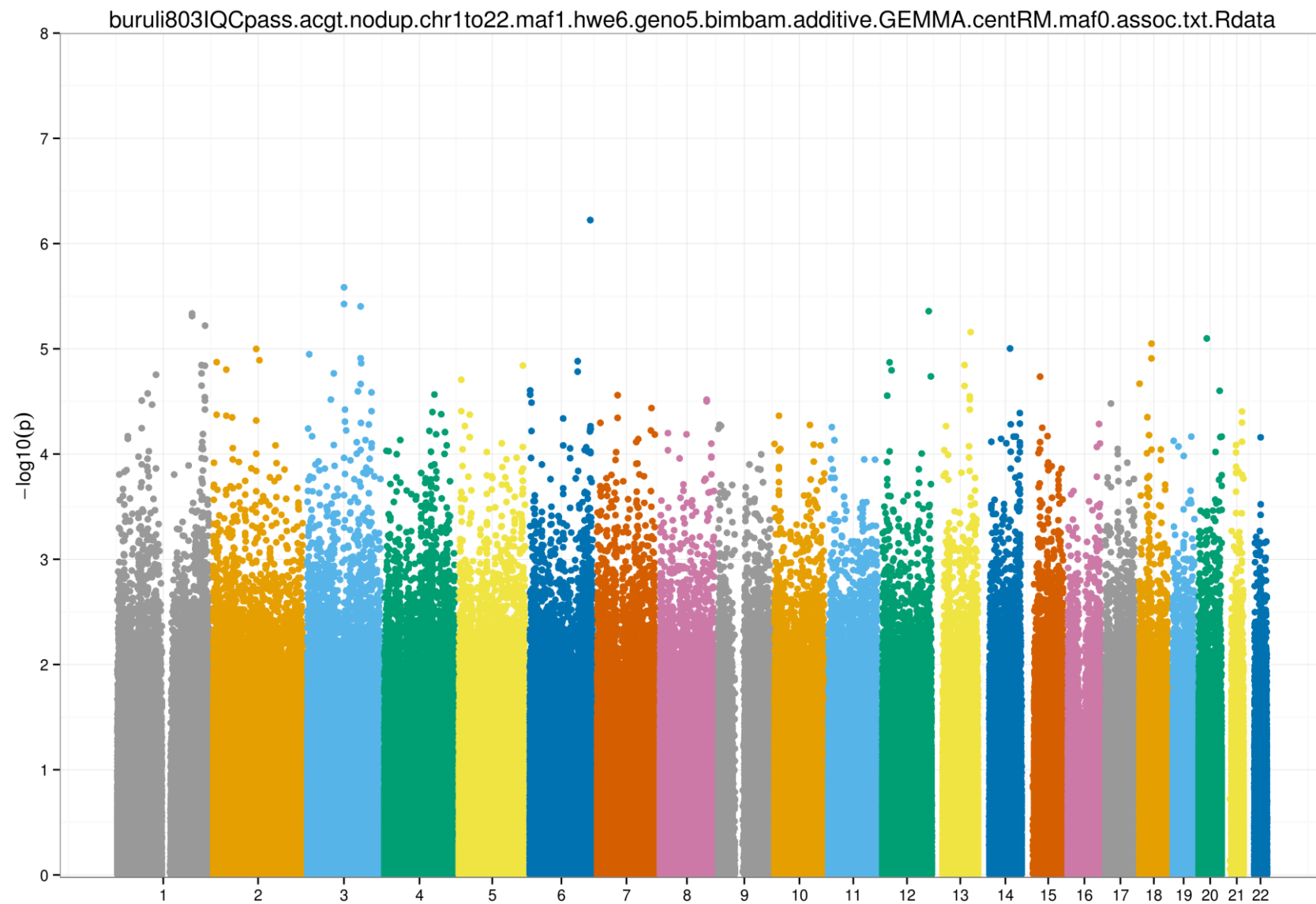


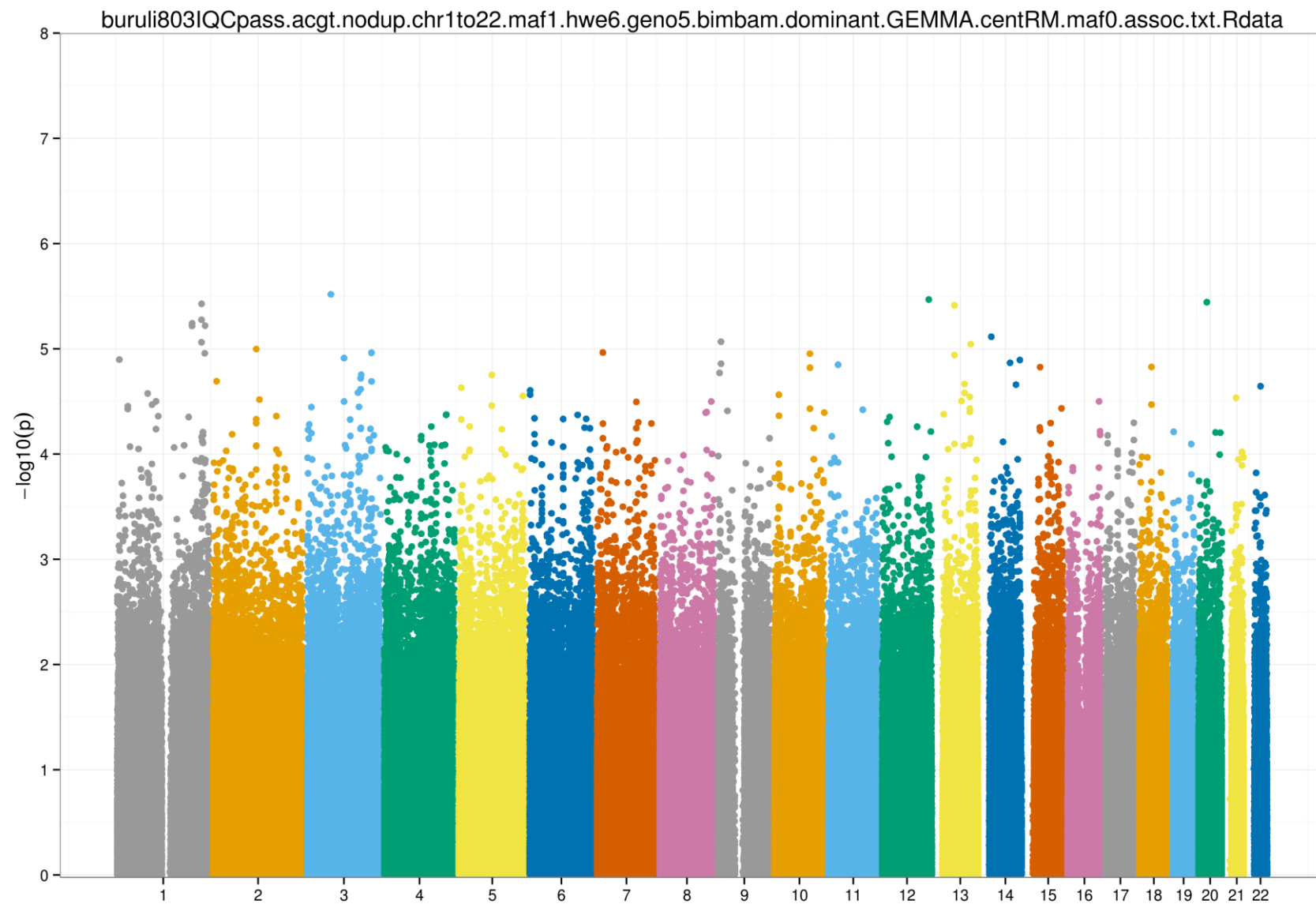


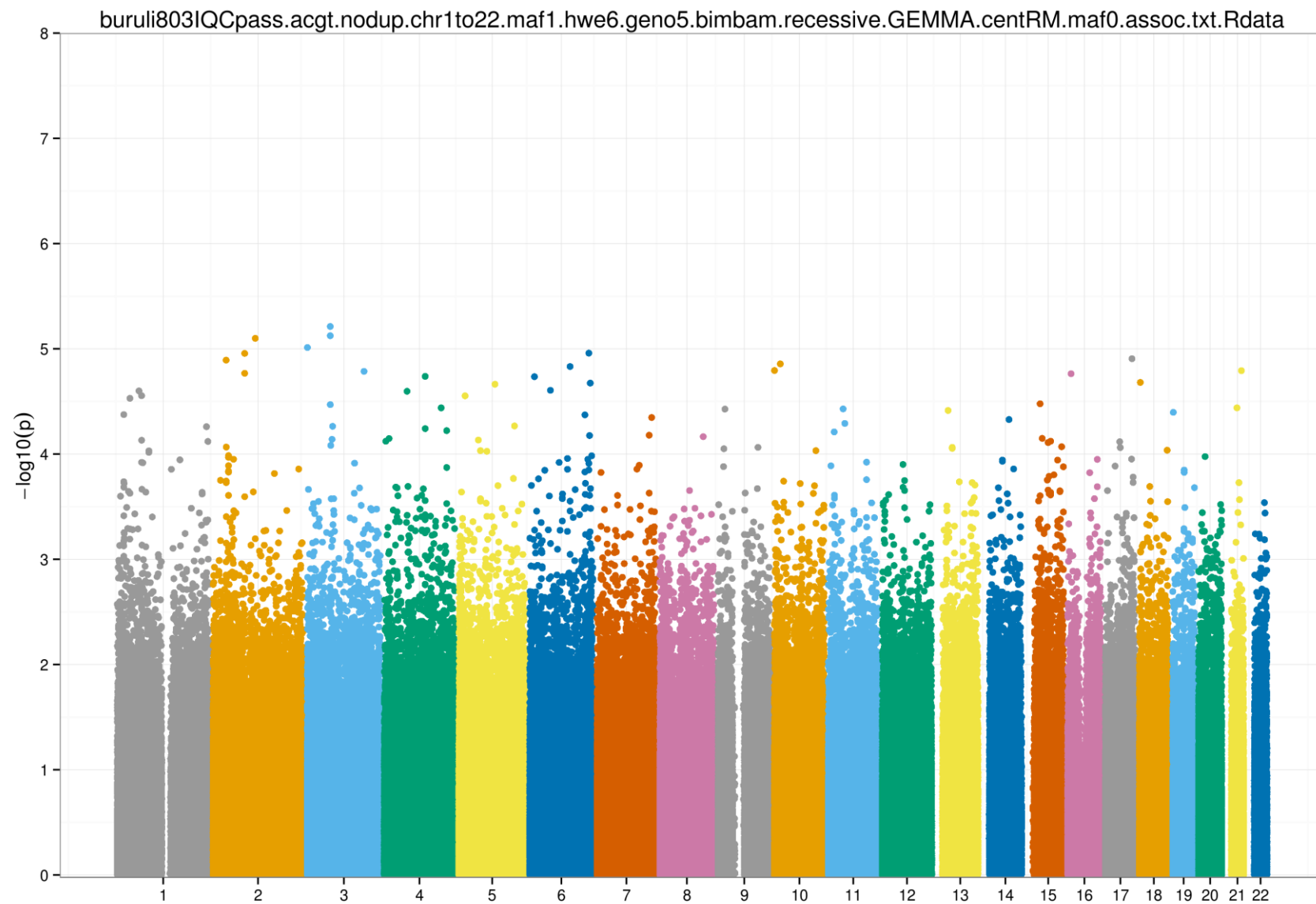


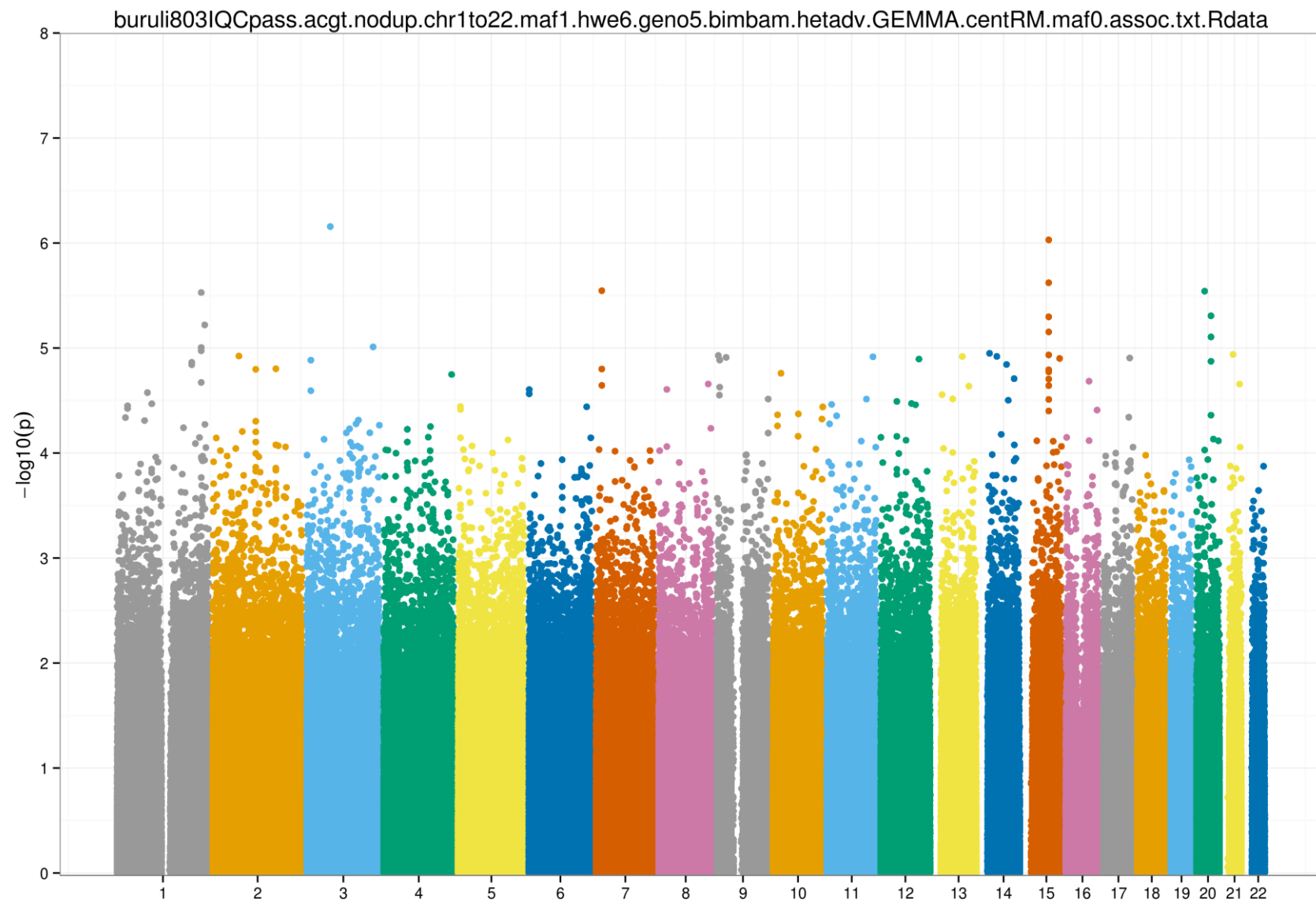


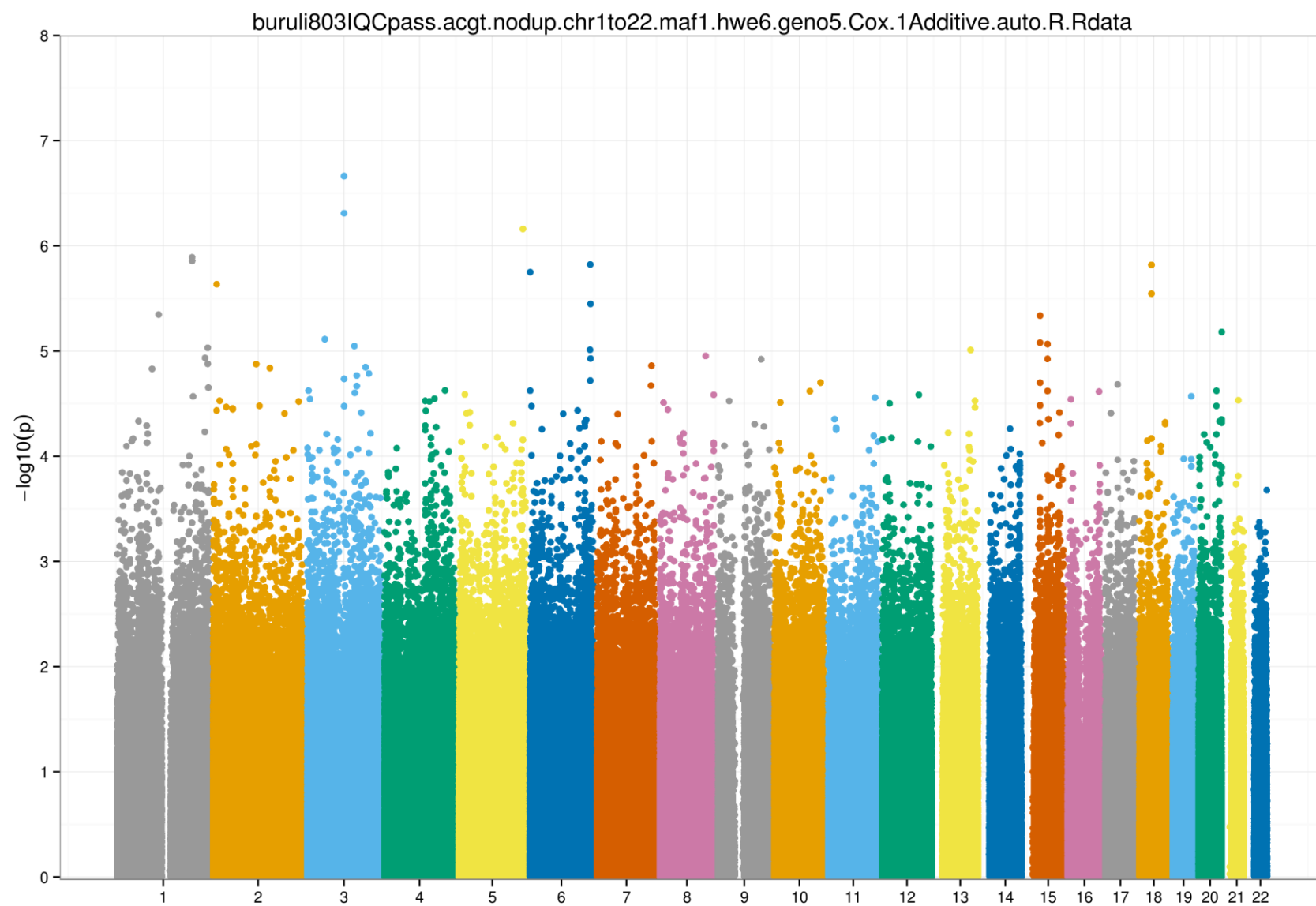




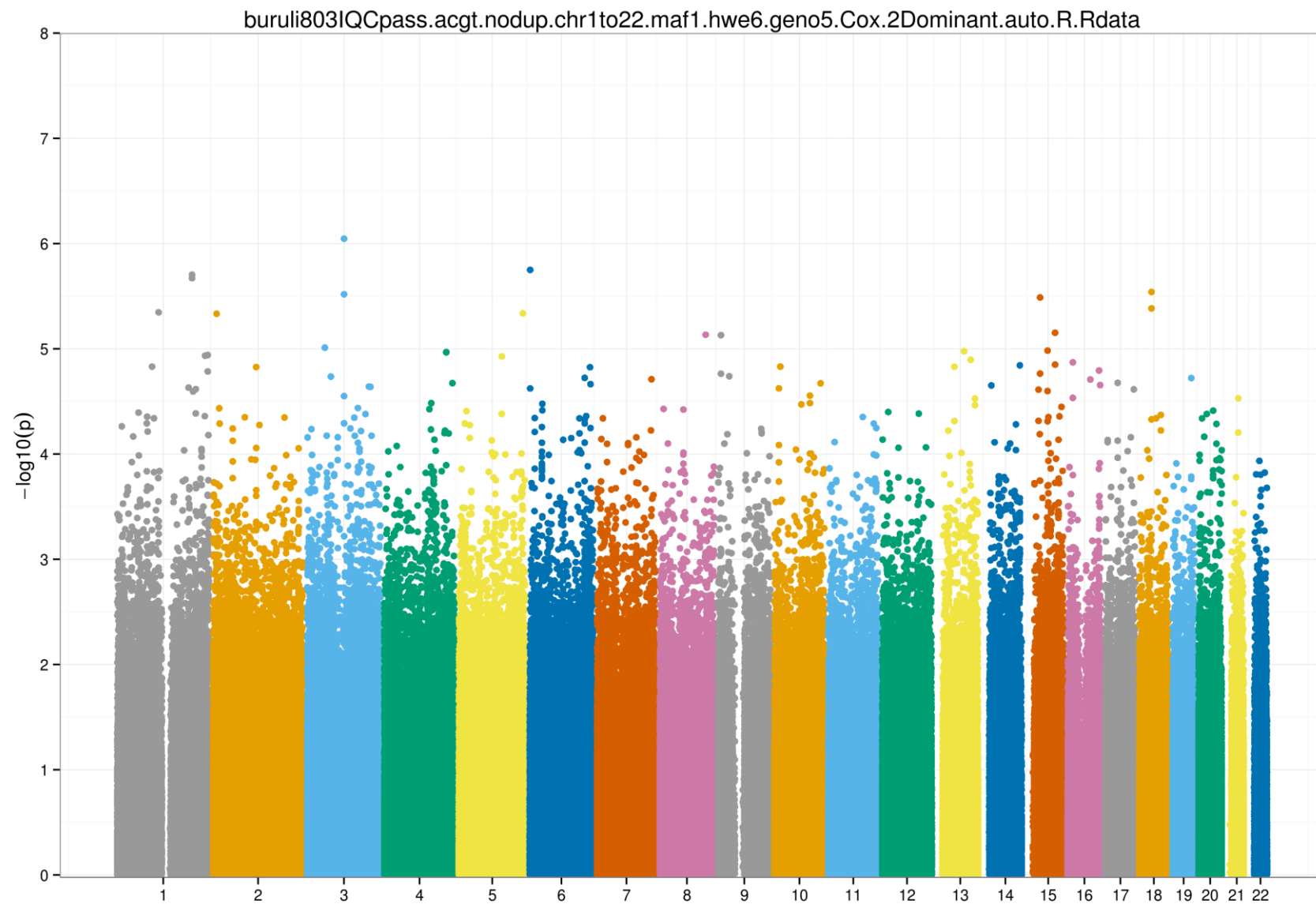


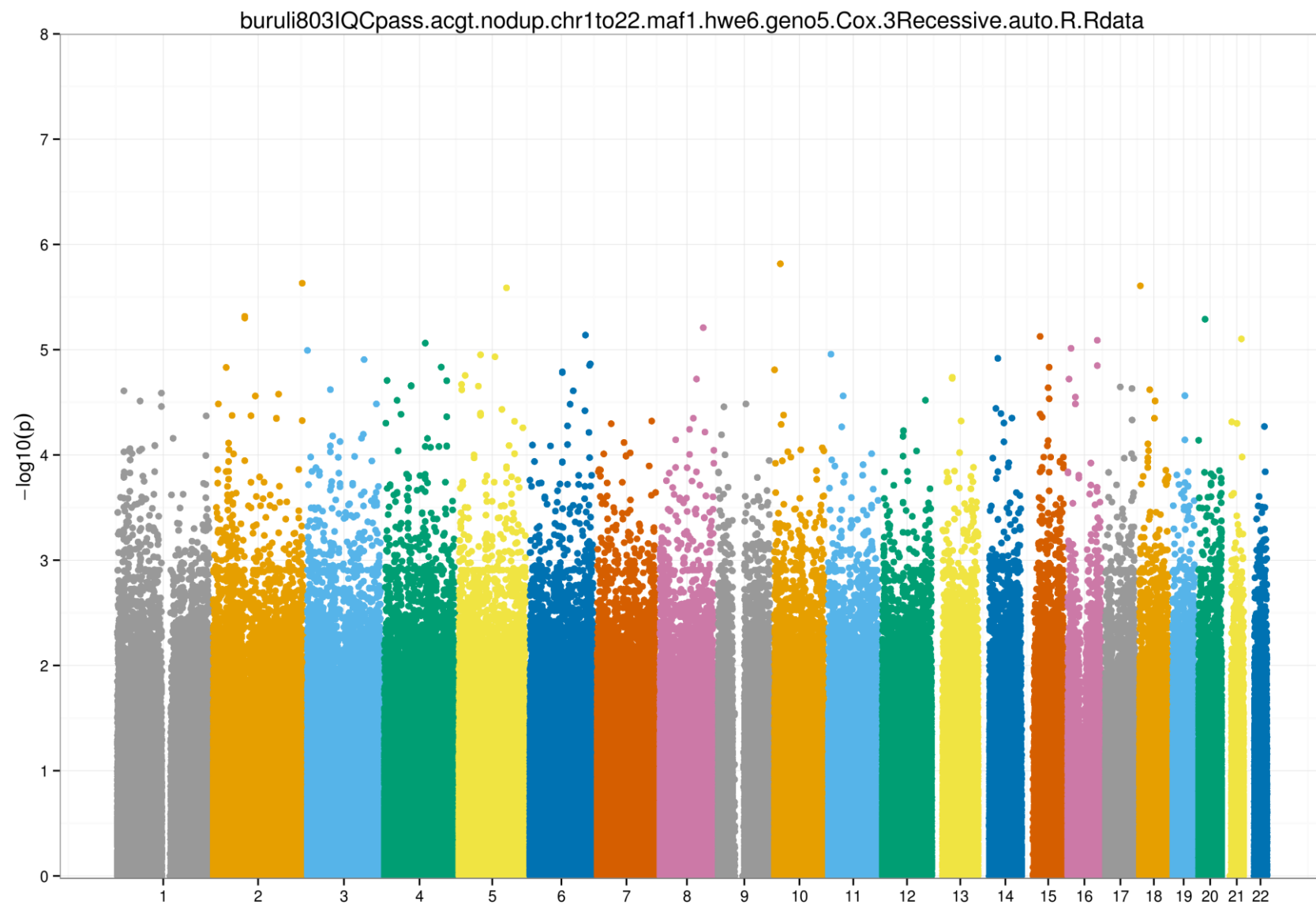


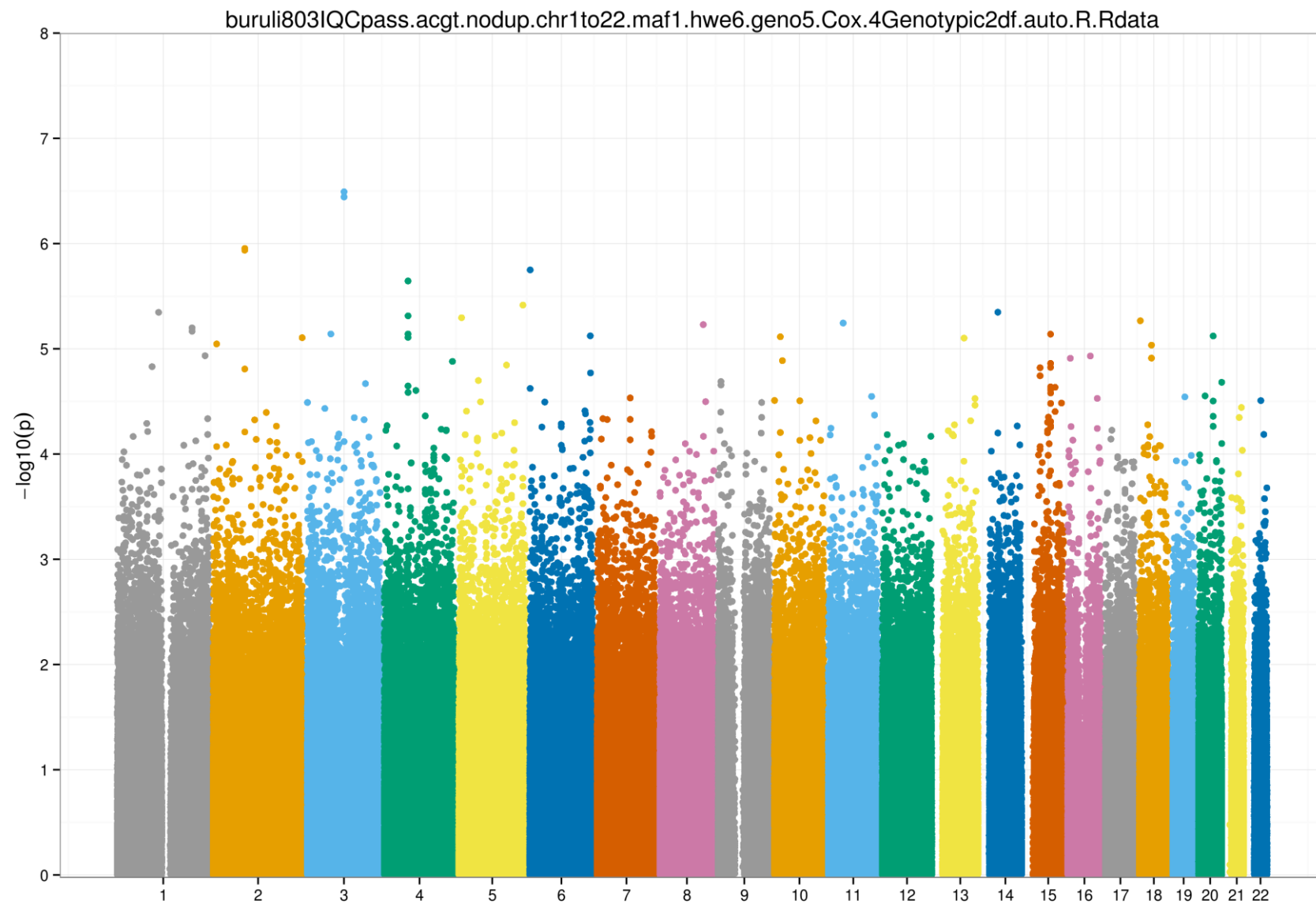


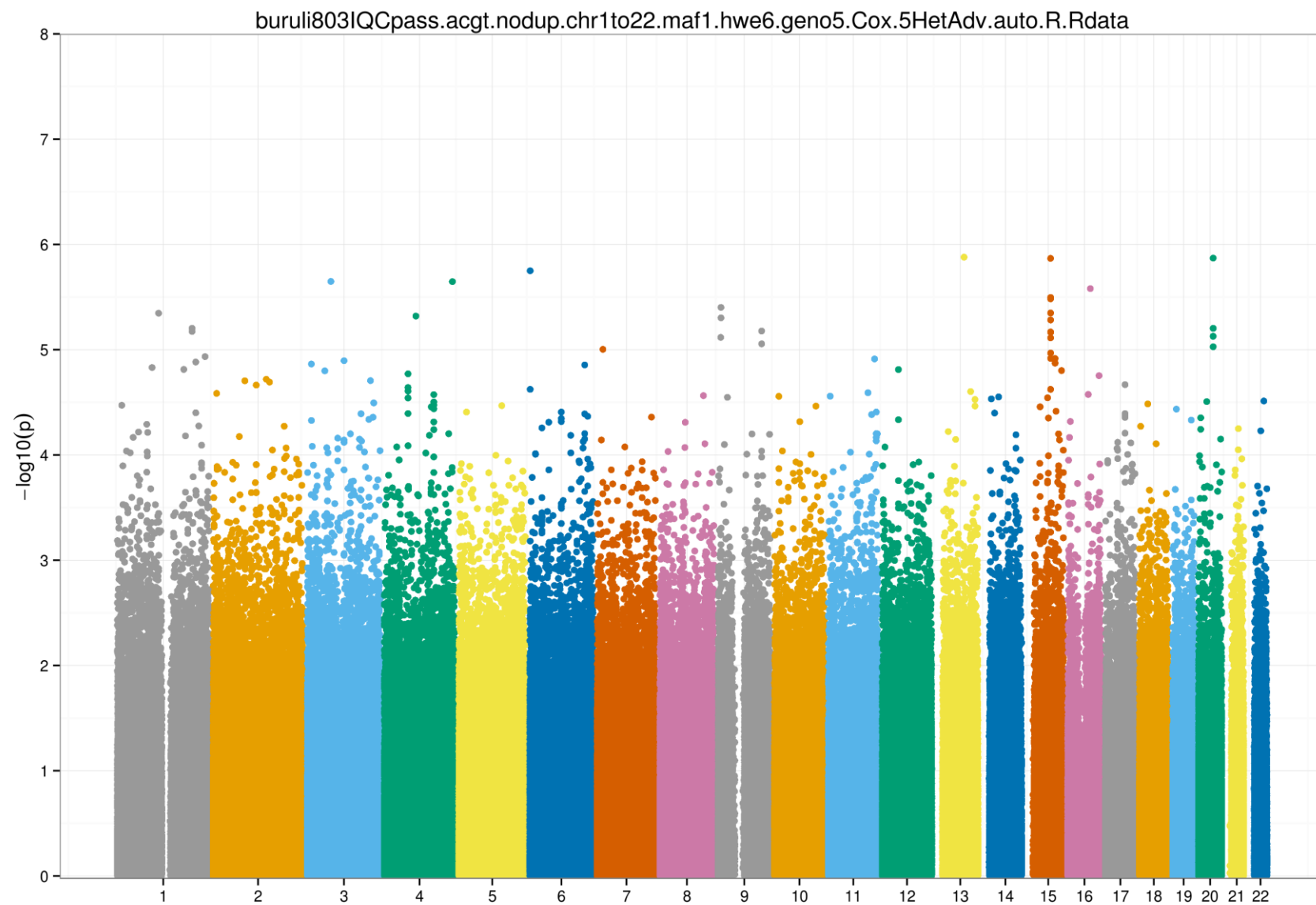


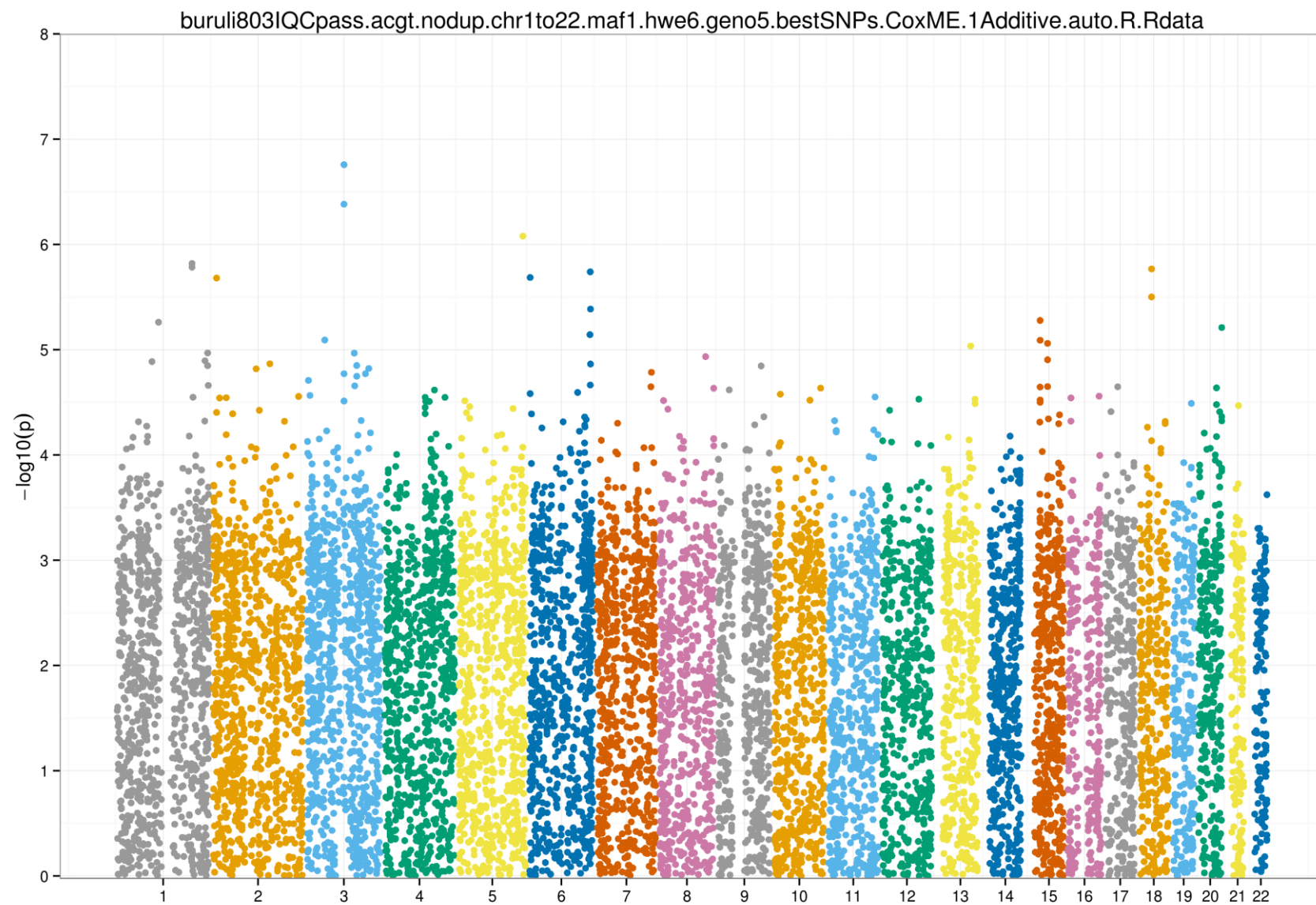


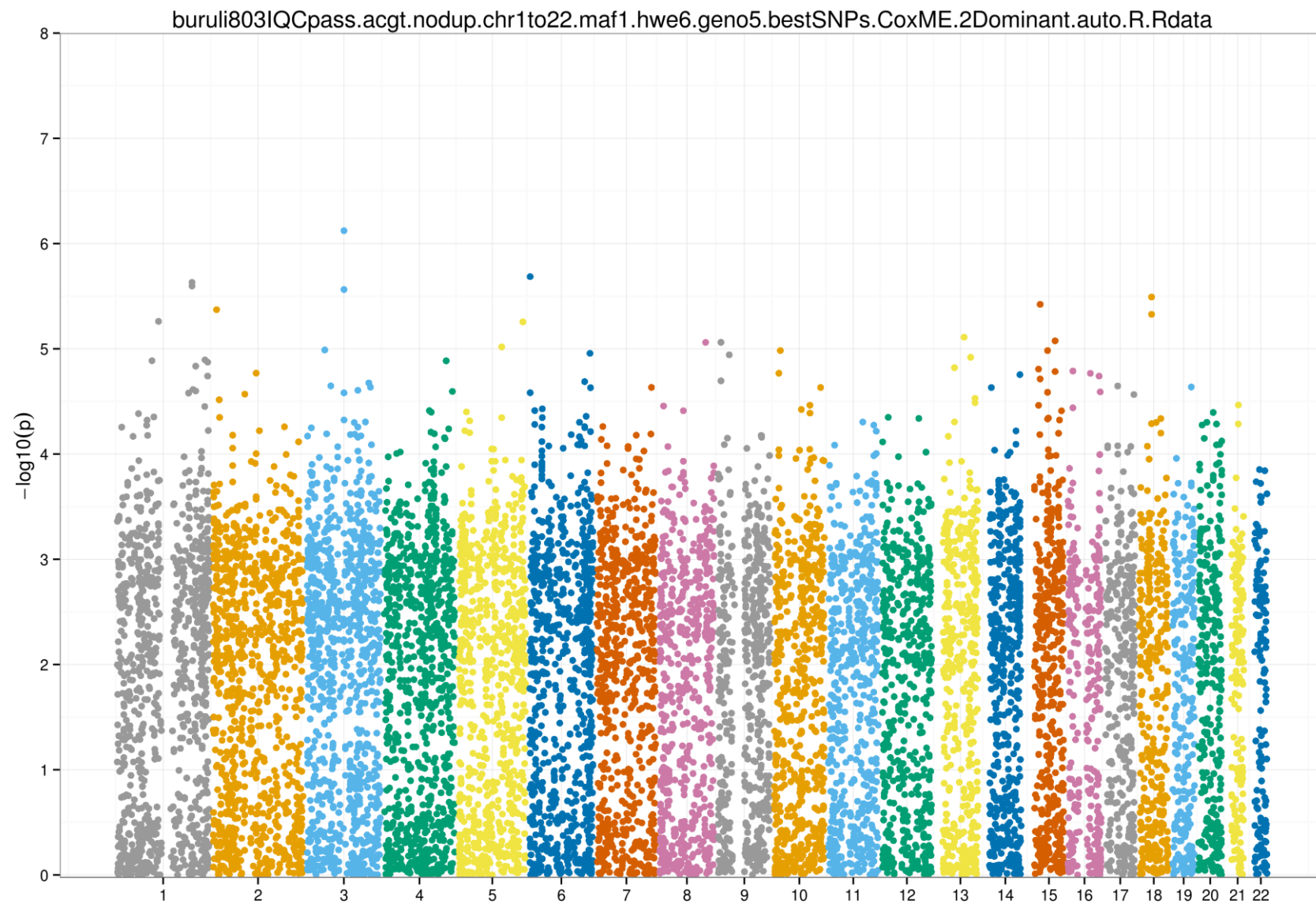




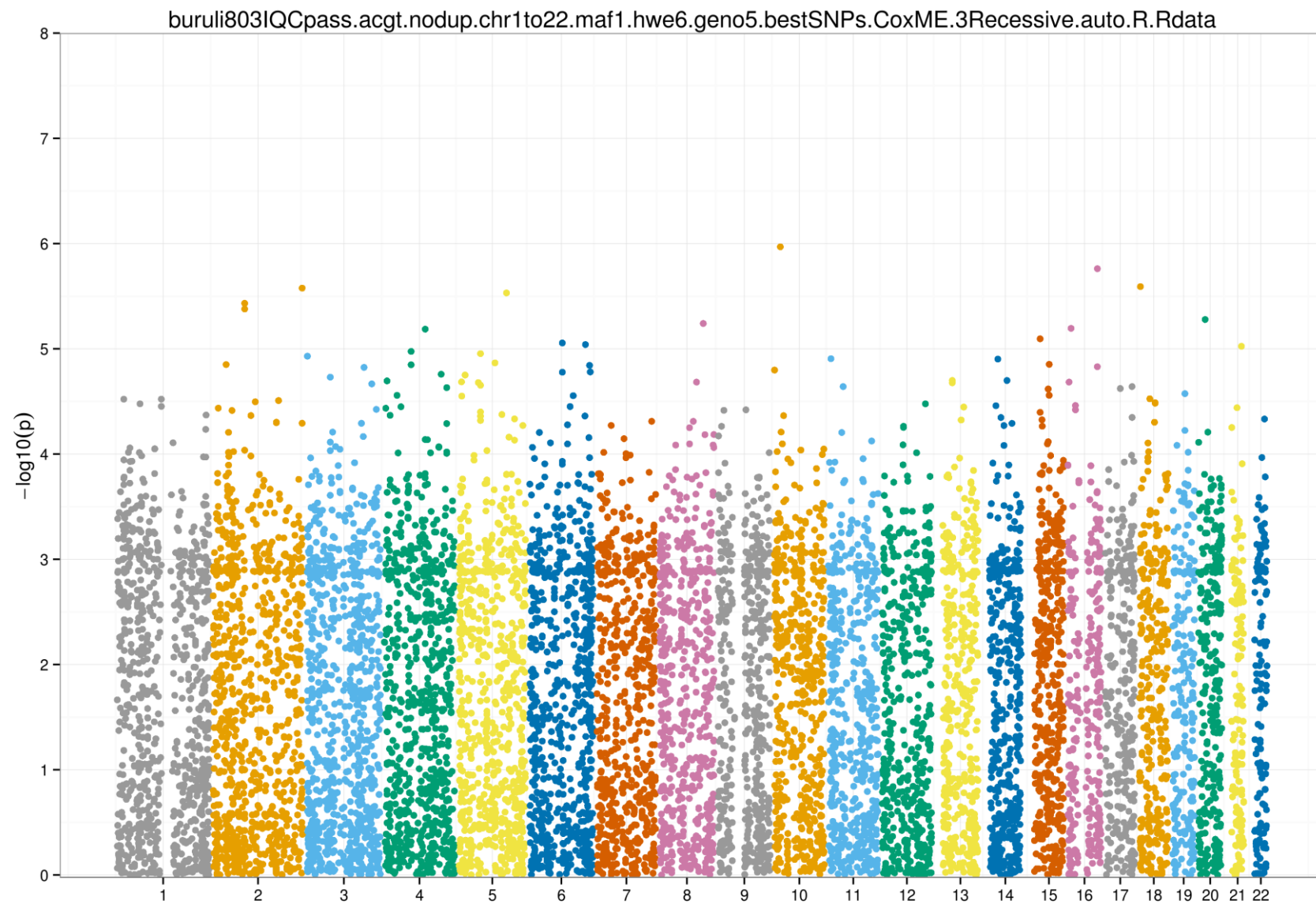






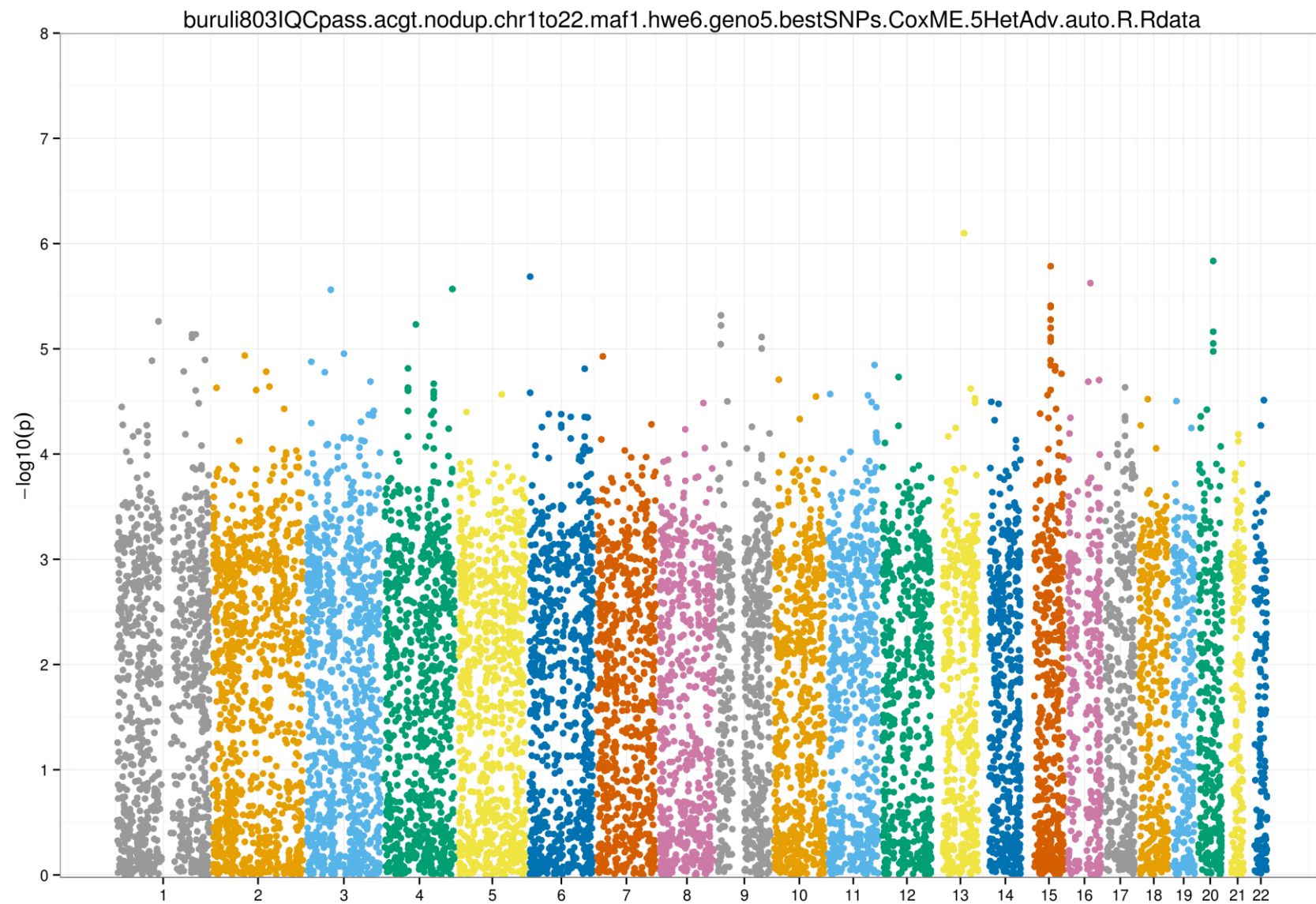












## **GENETIQUE COMPLEXE DE L'ULCERE DE BURULI :**

### **DETAIL DES MANHATTAN PLOTS**

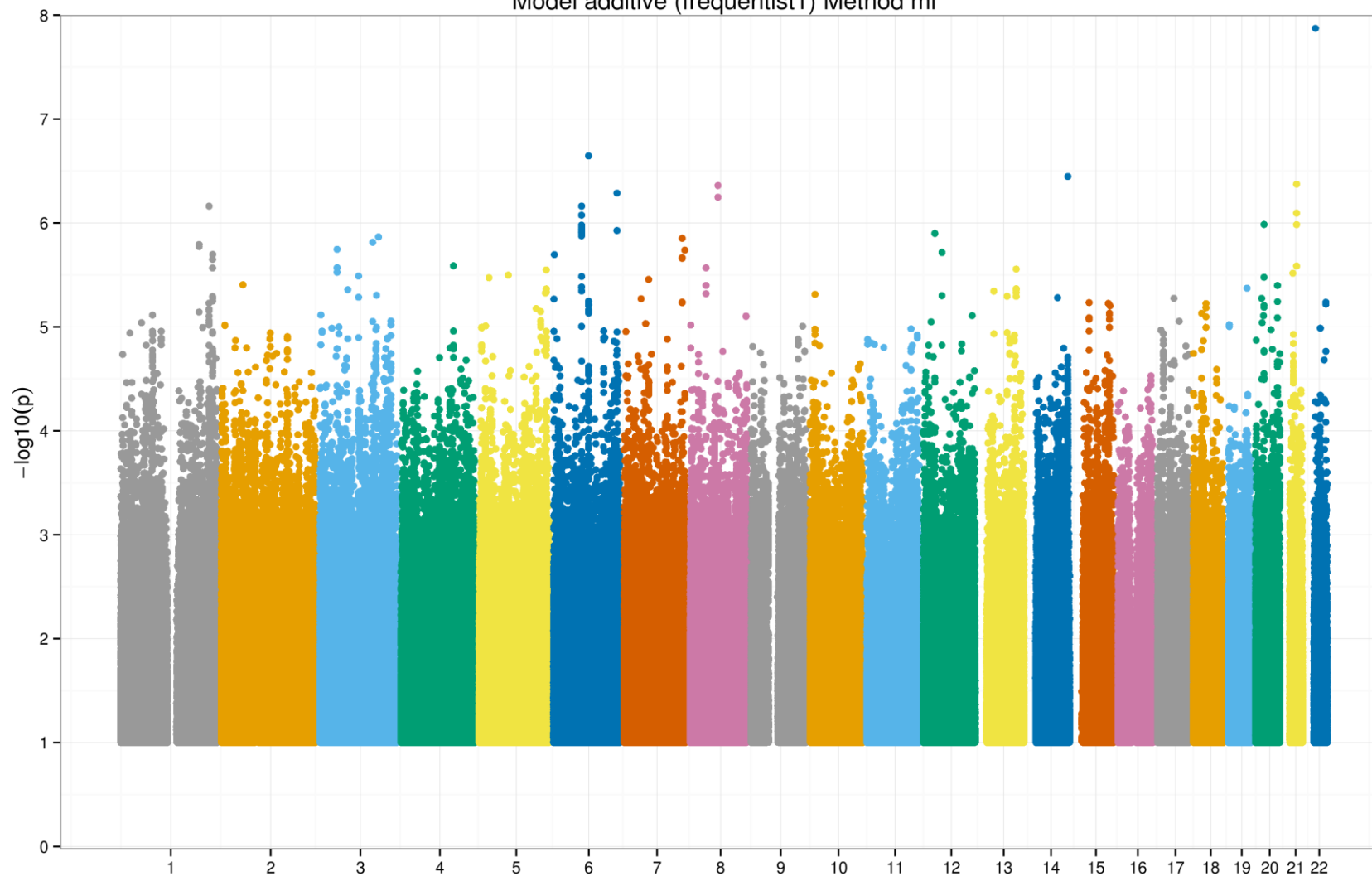
#### **SNPs IMPUTES**

Les SNPs imputés ont été analysés via SNPtestv2.4 par régression logistique prenant en compte l'incertitude de l'imputation génotypique par la méthode ML. Sont successivement représentés les modèles génétiques additifs, dominants, récessifs, génotypiques (à 2 degrés de liberté), avantage aux hétérozygotes.

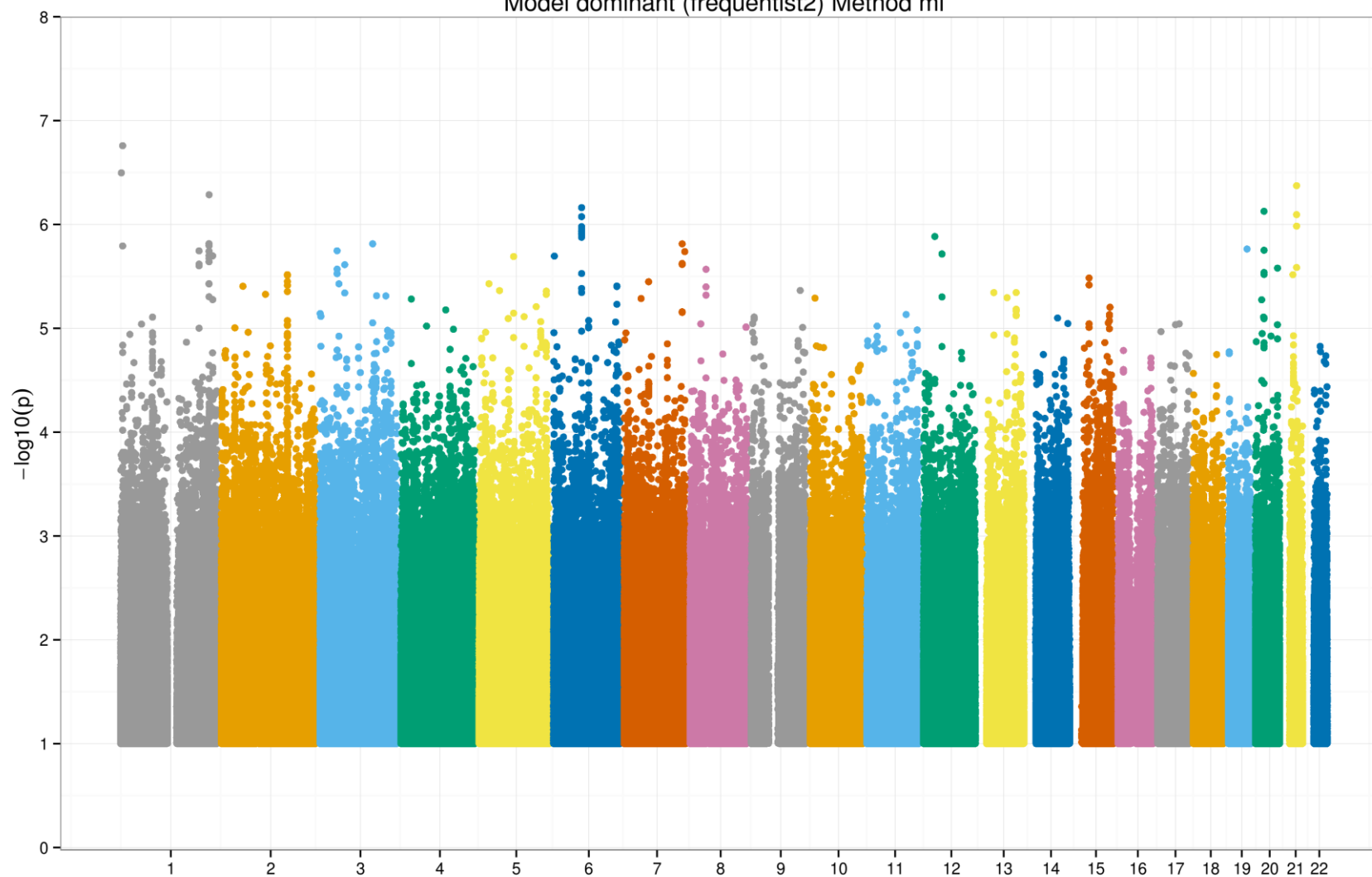
Tous les Manhattan plots représentent l'ensemble des SNPs imputés non monomorphes avec une information supérieure à 0.6, soit de l'ordre de 20 millions de SNPs. Pour accélérer la production de l'image les SNPs avec une pvalue supérieure à 0.10 n'ont pas été représentés.

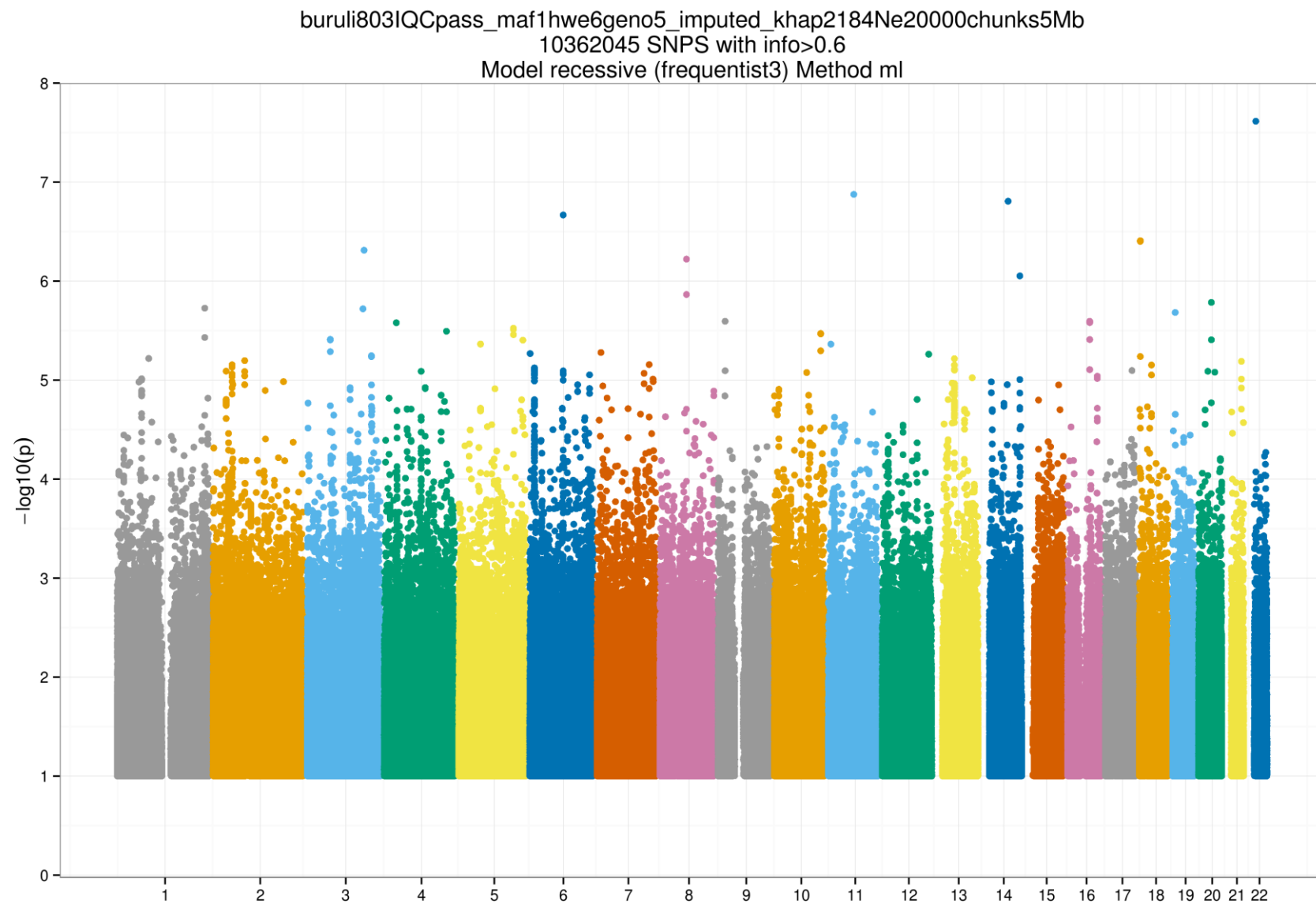


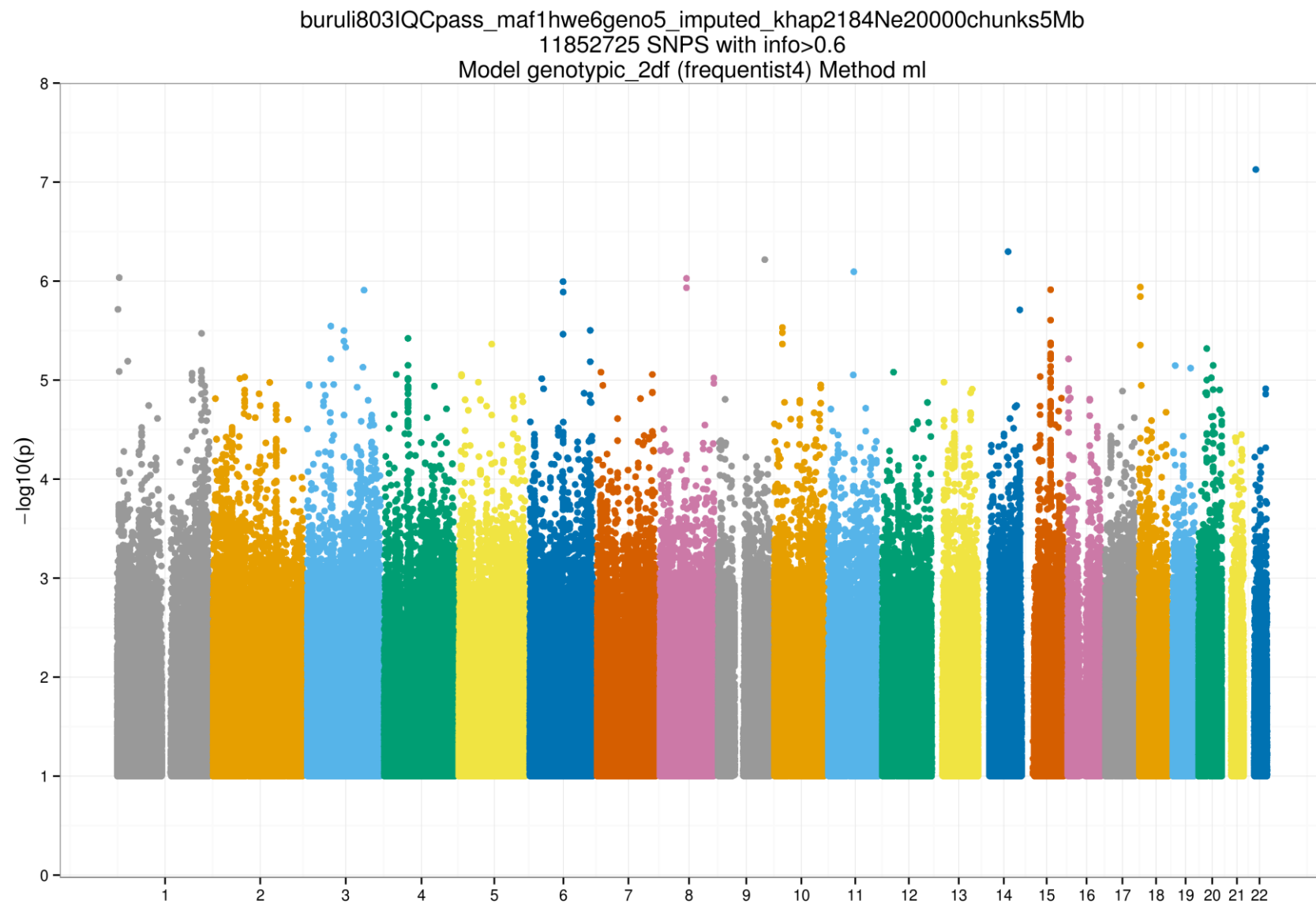
buruli803IQcpass\_maf1hwe6geno5\_imputed\_khap2184Ne20000chunks5Mb  
18703045 SNPS with info>0.6  
Model additive (frequentist1) Method ml



buruli803IQCpass\_maf1hwe6geno5\_imputed\_khap2184Ne20000chunks5Mb  
18549458 SNPS with info>0.6  
Model dominant (frequentist2) Method ml



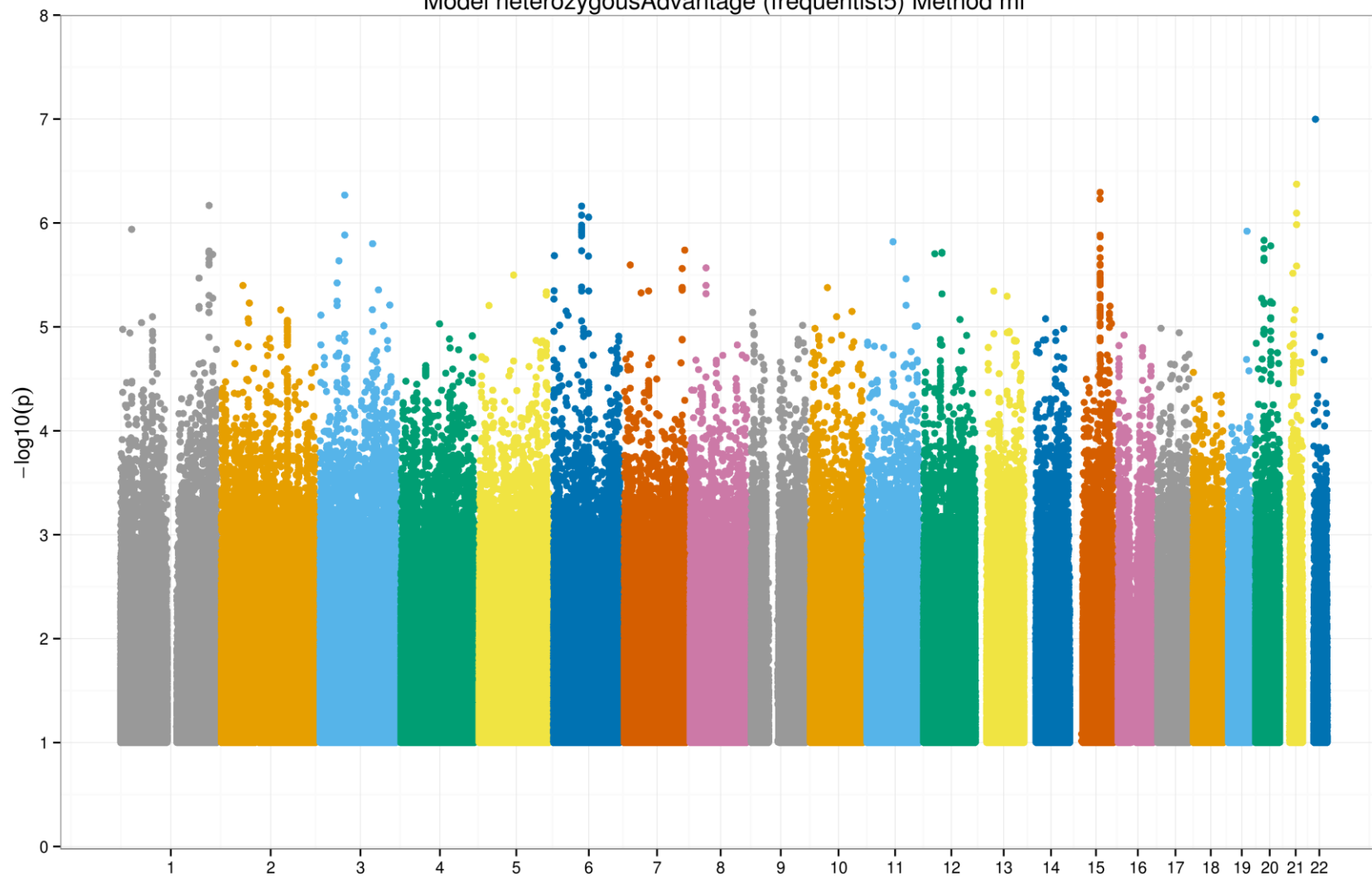




buruli803IQCpass\_maf1hwe6geno5\_imputed\_khap2184Ne20000chunks5Mb

18699300 SNPS with info>0.6

Model heterozygousAdvantage (frequentist5) Method ml







## **RESUME : Epidémiologie et Génétique Humaine de l'ulcère de Buruli**

L'ulcère de Buruli (UB), infection à *Mycobacterium ulcerans*, troisième mycobactériose mondiale, connaît une émergence rapide depuis 1980, essentiellement dans les pays d'Afrique subsaharienne. Nous avons constitué la plus grande cohorte de cas confirmés à ce jour rassemblant plus de 1200 patients traités au CDTUB de Pobè au Bénin entre 2005 et 2011, afin de décrire l'épidémiologie classique de la maladie et d'explorer l'architecture génétique de la susceptibilité à cette maladie.

Les patients atteints d'UB sont des enfants (âge médian au diagnostic de 12 ans), présentant une lésion unique (96%), large (plus de 15 cm, 36%), ulcérateuse (66%) du membre inférieur (60%). Nous rapportons une présentation clinique atypique de l'UB, dans laquelle les patients présentent exclusivement une ostéomyélite à *M. ulcerans*. Le sex-ratio varie avec l'âge : les garçons sont majoritaires parmi les enfants (57% de patients masculins chez les moins de 15 ans), et les femmes parmi les adultes (33% de patients masculins). La présentation clinique dépend de l'âge et du sexe. 9% des patients masculins ont présenté une ostéomyélite contre 4% des patients féminins. Un an après la fin du traitement, 22% des patients présentent des séquelles fonctionnelles fixées. Une présentation clinique comportant une lésion oedémateuse, osseuse, de grande taille ou plusieurs lésions est significativement associée avec le développement de séquelles fonctionnelles (OR 7.64, IC95% [5.29-11.31]). Les patients coinfectés par le VIH ont un risque significativement plus élevé de développer un UB sévère (OR 2.77, IC95% [1.32-6.33]).

Nous avons exploré l'architecture génétique de la susceptibilité à l'UB dans une perspective mendélienne et une perspective complexe. Le cas le plus sévère de la maladie observé dans ce centre appartient à une famille consanguine dans laquelle la ségrégation du phénotype suggère un défaut génétique mendélien récessif. Une analyse de liaison génétique par cartographie d'homozygotie suggère l'implication du locus des bêta-défensines sur le chromosome 8 dans la pathogénèse de l'UB, et mène à l'identification d'une délétion homozygote ségrégeant parfaitement avec la maladie. Dans une perspective complexe, une étude d'association pangénomique a été réalisée après génotypage d'une cohorte de 400 cas et 400 témoins exposés sur plus de 2 millions de SNPs par la puce Illumina Omni2.5 et a permis l'identification de nombreux signaux d'intérêt. L'étude de répllication est en cours.

La compréhension de la physiopathologie de l'infection à *M. ulcerans* est cruciale pour générer de nouvelles pistes thérapeutiques et vaccinales. La dissection du contrôle génétique de l'infection par l'hôte est en ce sens indispensable.

## **DISCIPLINE-SPECIALITE DOCTORALE**

Génétique Statistique

## **MOTS-CLES**

Maladie infectieuse, Epidémiologie, Génétique Humaine Mendélienne et Complexe, Mycobactérie, *Mycobacterium ulcerans*, Ulcère de Buruli, Bénin, Analyse de liaison, Défensine, Analyse d'association pangénomique, GWAS.

## **INTITULE ET ADRESSE DE L'UFR**

Faculté de Médecine Paris Descartes  
15, rue de l'Ecole de Médecine,  
75270 Paris Cedex 06